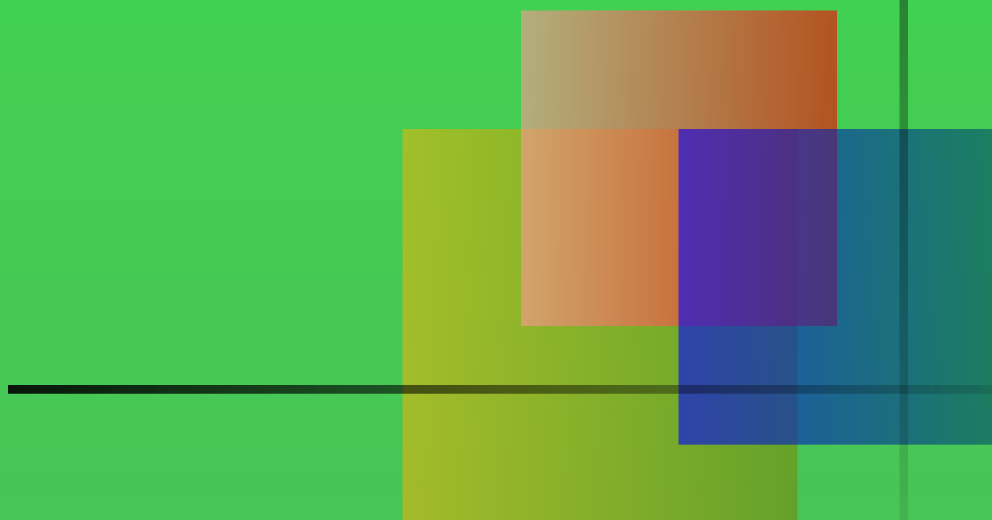
# Álgebra lineal

## con aplicaciones y Python

*Ernesto Aranda*



# Álgebra lineal con aplicaciones y Python

Ernesto Aranda

Título: *Álgebra lineal con aplicaciones y Python*
*Primera Edición, 2013*

CC BY Ernesto Aranda Ortega, 2013

Impreso por Lulu.com

Composición realizada con LaTeX

Todas las imágenes del libro han sido realizadas por el autor a excepción de las figuras 2.1 y 2.3b, debidas a Alain Matthes y la figura 2.4a de Bogdan Giuşcă

Este libro está disponible en descarga gratuita en la dirección
http://matematicas.uclm.es/earanda/?page_id=152

"Aprender matemáticas es un proceso de aprender
a hacer algo, no de adquirir conocimientos."

*J.M. Sanz-Serna*

**Diez lecciones de Cálculo Numérico**[1]

[1] Universidad de Valladolid, 1998.

# Prólogo

La palabra álgebra proviene del término árabe *yabr* que significa “reducción” y aparece por primera vez en el tratado del matemático persa Muhammad ibn Musa al-Jwarizmi titulado *Kitab al-yabr wa-l-muqabala* (“Compendio de cálculo por el método de completado y balanceado”) y dedicado especialmente a la solución de ecuaciones (lineales y cuadráticas). Es por ello que, a lo largo de la historia, el principal objetivo del Álgebra haya sido la resolución de ecuaciones. Sin embargo, en el s. XIX comienza a aparecer una temática transversal que se alimenta de problemas provenientes de la geometría, el análisis, la teoría de números y por supuesto, la teoría de ecuaciones, que desemboca en el estudio de estructuras matemáticas abstractas conformando lo que hoy en día se conoce como *álgebra moderna*. Este texto está dedicado esencialmente al estudio de una de tales estructuras abstractas, los *espacios vectoriales*, dentro de lo que se conoce como *álgebra lineal*, y en el que los *sistemas de ecuaciones lineales* juegan un papel central.

La división temática de este texto comprende los contenidos correspondientes a la asignatura de Álgebra de los grados de ingeniería de la Universidad de Castilla-La Mancha, en los que el autor imparte docencia desde hace años, aunque el material que presentamos puede ser también una referencia útil en carreras científico-técnicas en las que es habitual una formación en álgebra lineal, al constituir ésta una herramienta matemática básica en numerosas disciplinas.

En lo que se refiere a los contenidos del texto, habría que dividir el libro en dos partes: en los tres primeros temas que tratan sobre *números complejos*, *matrices y determinantes* y *sistemas de ecuaciones lineales* presentamos las herramientas esenciales que conforman el soporte básico del cual se nutren el resto de temas. Aunque es probable que el lector haya tenido contacto con estos conceptos en cursos anteriores, seguramente encontrará que el tratamiento de los mismos y la notación empleada no le son tan habituales. Sin embargo hemos de resaltar la importancia que supone entender y manejar apropiadamente el lenguaje matemático. Por ello hemos incluído en un primer apéndice (apéndice A) una serie de *conceptos generales* en el que tratamos de familiarizar al lector con la notación y el uso de sentencias lógicas de una forma intuitiva, a la vez que introducimos unas cuantas nociones de teoría de conjuntos, funciones y estructuras algebraicas. El tratamiento en este apéndice dista mucho de ser matemáticamente riguroso y solo pretende fijar algunas ideas básicas en el lector.

Aunque aparece al final del texto, recomendamos encarecidamente la lectura del mismo antes de abordar los demás temas.

En una segunda parte, que comenzaría con el tema 4, se desarrolla el material típico en un curso de álgebra lineal: *espacios vectoriales*, *aplicaciones lineales*, *diagonalización* y *espacios euclídeos*. Hemos incluído además un tema dedicado al estudio de *ecuaciones lineales en diferencias* y otro al *espacio afín*, con los que cubrimos los contenidos especificados en los descriptores de la asignatura para el primer curso de los nuevos grados de ingeniería.

En cada uno de los temas se pretende ofrecer al lector varias perspectivas de la materia. Así, se ha proporcionando una relación de definiciones y resultados con un enfoque muy matemático, incluyendo una buena cantidad de demostraciones que pueden ser omitidas en cursos de corte más técnico, a la vez que hemos tratado de ilustrar estos resultados con numerosos ejemplos. Por otro lado, hemos incluído en cada tema alguna aplicación relevante que pone de manifiesto que el álgebra no solo no es una disciplina abstracta, sino que por el contrario, sus herramientas juegan un papel destacado en diversas aplicaciones. Y además, teniendo en cuenta la realidad actual en la que los ordenadores están cada vez más presentes en todos los contextos, hemos tratado de ilustrar el uso del lenguaje de programación Python para llevar a cabo buena parte de los cálculos involucrados en cada uno de los temas.

Es importante señalar que éste no es un libro para aprender a programar en Python pues no hace un tratamiento profundo de una serie importante de características del lenguaje. No obstante, con las indicaciones incluídas en cada tema, creemos que el lector podrá usar el lenguaje para el propósito que planteamos aquí, que no es más que ayudar en los cálculos, en ocasiones tediosos, que son necesarios realizar a lo largo del texto. En cualquier caso incluímos en un segundo apéndice (apéndice B) una breve introducción de los aspectos esenciales para el manejo de Python, de obligada lectura para los profanos en el lenguaje.

Como en cualquier otra asignatura de matemáticas, el aspecto esencial que se persigue es la adquisición de destrezas más que conocimientos. Tal y como reza la cita con la que abrimos este texto: *aprender matemáticas es un proceso de aprender a hacer algo, no de adquirir conocimientos*. Obviamente no debe interpretarse esto como una invitación a dejar de lado los contenidos teóricos de la asignatura, por otra parte imprescindibles. Pero es fundamental tener presente que **el aprendizaje debe ir encaminado a la resolución de problemas**. No es sino mediante los ejercicios como el lector podrá poner a prueba el éxito de su aprendizaje y es por ello que recomendamos encarecidamente que sea ésa la labor central de su trabajo. El estudio de la teoría es simplemente una condición necesaria para poder resolver los problemas.

En ese sentido hemos incluído en cada tema una sección final con ejercicios de diversa naturaleza. Por una parte aparecen *ejercicios de repaso* con los que se pretende que el lector pueda comprobar la correcta adquisición de conocimientos y herramientas básicas de cada tema. El apartado de *problemas* comprende ejercicios más variados en los que se requiere ir un paso más allá que en los

ejercicios de repaso, bien precisando del uso simultáneo de varios conceptos o de algunas técnicas más depuradas. Hay también un apartado dedicado a *ejercicios teóricos* de naturaleza más matemática y con los que se persigue que el lector ponga en práctica técnicas similares a las aprendidas en las demostraciones de los resultados, y finalmente, hay un breve apartado de *ejercicios adicionales* de carácter opcional que tienen que ver con las aplicaciones y/o el uso de Python que reservamos para el lector interesado. Hemos marcado con * aquellos ejercicios que pueden resultar de mayor dificultad y hemos incluído en el apéndice C las soluciones a los ejercicios de repaso.

La versión electrónica de este libro estará siempre disponible en descarga directa en la dirección que figura al final de la página de créditos. Agradecemos de antemano al lector cualquier sugerencia que nos haga llegar para mejorar el contenido de este libro.

Ciudad Real, 28 de enero de 2013.
El autor.

# Índice general

Prólogo . . . . . 3

1 **Números complejos** . . . . . 11

1.1 El cuerpo de los números complejos . . . . . 11
1.2 Representación gráfica: módulo y argumento . . . . . 15
1.3 Forma trigonométrica y forma polar . . . . . 18
1.4 Potencia y raíz n-ésima de un número complejo . . . . . 20
1.5 Cálculo con Python . . . . . 22
1.6 Una breve incursión en el mundo fractal . . . . . 26
1.7 Ejercicios . . . . . 32

2 **Matrices y determinantes** . . . . . 35

2.1 Matrices: primeras definiciones . . . . . 35
2.2 Inversa de una matriz . . . . . 45
2.3 Sistemas de ecuaciones lineales . . . . . 47
2.4 Cálculo de la inversa mediante operaciones elementales . . . . . 54
2.5 Determinante de una matriz . . . . . 58
2.6 Rango de una matriz . . . . . 69
2.7 Aplicación de los determinantes al cálculo de la inversa . . . . . 74
2.8 Cálculo con Python . . . . . 76
2.9 Breve introducción a la Teoría de Grafos . . . . . 80
2.10 Ejercicios . . . . . 84

**3 Sistemas de ecuaciones lineales** ..... 91

3.1 Sistemas de ecuaciones lineales: primeras definiciones ..... 91
3.2 Teorema de Rouché-Frobenius ..... 98
3.3 Álgebra Lineal Numérica ..... 102
3.4 Cálculo con Python ..... 119
3.5 Aplicación: resolución de ecuaciones diferenciales ..... 124
3.6 Ejercicios ..... 129

**4 Espacios vectoriales** ..... 135

4.1 La estructura de espacio vectorial ..... 135
4.2 Independencia lineal ..... 140
4.3 Bases y dimensión de un espacio vectorial ..... 146
4.4 Cambios de base ..... 157
4.5 Subespacios vectoriales ..... 162
4.6 Cálculo con Python ..... 179
4.7 Aplicación: *Lights Out!*, primera parte ..... 182
4.8 Ejercicios ..... 185

**5 Aplicaciones lineales** ..... 191

5.1 Aplicaciones lineales entre espacios vectoriales ..... 191
5.2 Matriz de una aplicación lineal. ..... 195
5.3 Operaciones entre aplicaciones lineales ..... 200
5.4 Cambio de base en una aplicación lineal ..... 204
5.5 Núcleo y rango de una aplicación lineal ..... 209
5.6 Cálculos con Python ..... 214
5.7 Aplicación a la Criptografía ..... 215
5.8 Ejercicios ..... 221

**6 Diagonalización** ..... 227

6.1 Valores y vectores propias ..... 228
6.2 Polinomio característico ..... 234
6.3 Forma canónica de Jordan ..... 245

6.4 Cálculo con Python . . . . . 270
6.5 Aplicación: osciladores acoplados . . . . . 276
6.6 Ejercicios . . . . . 278

7 Ecuaciones lineales en diferencias . . . . . 285

7.1 Introducción . . . . . 285
7.2 Ecuaciones y sistemas lineales de primer orden . . . . . 288
7.3 Ecuaciones de orden superior . . . . . 294
7.4 Cadenas de Markov . . . . . 299
7.5 Aplicación: modelos biológicos . . . . . 305
7.6 Ejercicios . . . . . 308

8 Espacio vectorial euclídeo . . . . . 311

8.1 Producto escalar . . . . . 311
8.2 Ortogonalidad . . . . . 317
8.3 Método de los mínimos cuadrados . . . . . 337
8.4 Cálculo con Python . . . . . 340
8.5 Aplicación: Lights Out!, segunda parte . . . . . 342
8.6 Ejercicios . . . . . 344

9 Espacio afín . . . . . 349

9.1 Espacio afín y espacio métrico . . . . . 349
9.2 Variedades afines . . . . . 353
9.3 Problemas métricos . . . . . 357
9.4 Aplicaciones afines . . . . . 362
9.5 Aplicación: movimientos rígidos . . . . . 365
9.6 Cálculo con Python . . . . . 367
9.7 Ejercicios . . . . . 370

A Conceptos generales . . . . . 375

A.1 Teoría de conjuntos . . . . . 375
A.2 Funciones . . . . . 377

A.3 Estructuras algebraicas . . . . . 379
A.4 Principio de inducción . . . . . 380

B Introducción a Python . . . . . 383
B.1 Instalación de Python . . . . . 384
B.2 Aspectos básicos del lenguaje . . . . . 386
B.3 Bucles y condicionales . . . . . 389
B.4 Módulos . . . . . 391

C Soluciones a los ejercicios de repaso . . . . . 399
C.1 Números complejos . . . . . 399
C.2 Matrices y determinantes . . . . . 400
C.3 Sistemas de ecuaciones lineales . . . . . 402
C.4 Espacios vectoriales . . . . . 404
C.5 Aplicaciones lineales . . . . . 405
C.6 Diagonalización . . . . . 407
C.7 Ecuaciones lineales en diferencias . . . . . 410
C.8 Espacio vectorial euclídeo . . . . . 411
C.9 Espacio afín . . . . . 412

Índice terminológico . . . . . 413

Índice de autores . . . . . 418

Bibliografía . . . . . 421

# 1 Números complejos

Este tema está dedicado a introducir un nuevo conjunto de números: los *números complejos*, habitualmente denotado como $\mathbb{C}$, y que será necesario usar en determinados momentos a lo largo de este texto. Se pretende dar una introducción rápida a las operaciones algebraicas con números complejos, su representación gráfica y el cálculo de potencias y raíces. Además veremos un resultado central, conocido como el *Teorema Fundamental del Álgebra*, que nos será de especial utilidad en el tema 6. Dedicaremos también una sección a la realización de cálculos con Python y veremos cómo los números complejos permiten definir objetos interesantes como los fractales.

## 1 1 EL CUERPO DE LOS NÚMEROS COMPLEJOS

El conjunto de números complejos surge de forma parecida al resto de conjuntos numéricos. El lector debe conocer que el conjunto de números enteros (denotado por $\mathbb{Z}$) se necesita para poder realizar la operación diferencia en el conjunto de números naturales ($\mathbb{N}$), mientras que el conjunto de números racionales ($\mathbb{Q}$) es introducido para permitir la división entre números enteros. En la misma línea, el conjunto de números complejos viene a cubrir la imposibilidad en el conjunto de números reales ($\mathbb{R}$) de obtener la raíz de un número negativo.[1]

Recordemos que la resolución de una ecuación de segundo grado en $\mathbb{R}$ depende del signo del discriminante

$$ax^2 + bx + c = 0 \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

de modo que:

(i) si $b^2 - 4ac > 0$ tenemos dos soluciones,

[1] El término *número complejo* fue introducido por el alemán Carl Friedrich Gauss hacia 1831 aunque ya en los trabajos de matemáticos griegos surge la primera referencia conocida a raíces cuadradas de números negativos como resultado de una imposible sección de una pirámide. Mucho más tarde, el matemático italiano Gerolamo Cardano se dio cuenta de que podían manejarse cantidades más generales que los números reales cuando intentaba encontrar una fórmula para resolver ecuaciones cúbicas, en torno al año 1540.

(II) si $b^2 - 4ac = 0$ hay una única solución (doble),

(III) si $b^2 - 4ac < 0$ no hay solución (real).

El conjunto de números complejos se introduce con la finalidad de evitar la discusión anterior. Para ello comenzamos con la siguiente definición:

**Definición 1.1**

Se define la *unidad imaginaria* como $i = \sqrt{-1}$, o de forma equivalente, $i^2 = -1$.

Con esta definición, si $b^2 - 4ac < 0 \Rightarrow -(b^2 - 4ac) > 0$ y

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{-1}\sqrt{-(b^2 - 4ac)}}{2a} = \frac{-b \pm i\sqrt{-b^2 + 4ac}}{2a}$$

Observemos que $i$ es un nuevo número,[2] aquel cuyo cuadrado es $-1$, que permite salvar la imposibilidad existente en $\mathbb{R}$ de realizar operaciones que involucren raíces negativas. A partir de este nuevo número, definimos los números complejos del siguiente modo:

**Definición 1.2**

Un *número complejo* es una expresión de la forma $z = \alpha + i\beta$, donde $\alpha$, $\beta \in \mathbb{R}$. $\alpha$ es denominada la *parte real*, $\mathrm{Re}(z)$, mientras que $\beta$ es la *parte compleja o imaginaria*, $\mathrm{Im}(z)$. Esta expresión se conoce como *forma binomial* o *cartesiana* del número complejo.

Así, dos números complejos serán iguales si sus partes reales e imaginarias respectivas son iguales.

**Nota 1.1**

Si $\beta = 0 \Rightarrow z = \alpha \in \mathbb{R}$; es decir, los números reales forman un subconjunto de los números complejos, aquellos cuya parte imaginaria es nula. Si $\alpha = 0$ el número complejo $z = i\beta$ se denomina *imaginario puro*.

[2] La notación se debe al suizo Leonhard Euler, que la introdujo en 1777, aunque la adopción por parte de Gauss en 1801 fue la que lo hizo un símbolo habitual. En ingeniería eléctrica, electrónica y áreas relacionados, la unidad imaginaria es a menudo escrita como $j$ para evitar la confusión con la intensidad de corriente eléctrica, tradicionalmente denotada por $i$.

### Operaciones con números complejos

La **suma**, **resta** y **multiplicación** de números complejos[3] se realiza siguiendo las operaciones algebraicas habituales para binomios, es decir,

$$(\alpha_1 + i\beta_1) \pm (\alpha_2 + i\beta_2) = (\alpha_1 \pm \alpha_2) + i(\beta_1 \pm \beta_2)$$

$$\begin{aligned}(\alpha_1 + i\beta_1)(\alpha_2 + i\beta_2) &= \alpha_1\alpha_2 + i\alpha_1\beta_2 + i\alpha_2\beta_1 + i^2\beta_1\beta_2 \\ &= (\alpha_1\alpha_2 - \beta_1\beta_2) + i(\alpha_1\beta_2 + \alpha_2\beta_1)\end{aligned}$$

Nótese cómo se ha usado que $i^2 = -1$.

**Ejemplo 1.1**

Realicemos algunas operaciones básicas:

$$\begin{array}{ll} (2+3i)+(5-2i)=7-i & (1+i)-(2+5i)=-1-4i \\ (3+2i)+(5-2i)=8 & (1+i)-(1+2i)=-i \\ (2+3i)(5-2i)=16+11i & (1+i)(2+5i)=-3+7i \\ (3+2i)(5-2i)=19+4i & (1+i)(1+2i)=-1+3i \end{array}$$

Para efectuar la división de números enteros necesitamos la siguiente definición:

**Definición 1.3**

Si $z = \alpha + i\beta \in \mathbb{C}$, se define el *conjugado* de $z$, y se denotará por $\overline{z}$, como el número complejo $\overline{z} = \alpha - i\beta$.

Es decir, el conjugado de un número complejo no es más que un nuevo número complejo que tiene la misma parte real, y su parte imaginaria es la opuesta. La operación de conjugación posee las siguientes propiedades:

**Proposición 1.1**

Si $z, w \in \mathbb{C}$ se verifica:

[3] Las reglas para estas operaciones fueron desarrolladas por el matemático italiano Rafael Bombelli en un tratado de Álgebra publicado en 1572.

(I) $\overline{\overline{z}} = z$.

(II) $\overline{z + w} = \overline{z} + \overline{w}$.

(III) $z = \overline{z} \Leftrightarrow z$ es real.

(IV) $\overline{z} = -z \Leftrightarrow z$ es imaginario puro.

(V) Si $z = \alpha + i\beta \neq 0$, entonces $z\overline{z} = \alpha^2 + \beta^2 \in \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$.

(VI) $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$

La demostración se deja como ejercicio al lector (ejercicio 19).

La **división** entre números complejos se basa en usar adecuadamente la propiedad (V). En concreto,

$$\frac{\alpha + i\beta}{\gamma + i\delta} = \frac{(\alpha + i\beta)\overline{(\gamma + i\delta)}}{(\gamma + i\delta)\overline{(\gamma + i\delta)}} = \frac{(\alpha\gamma + \beta\delta) + i(-\alpha\delta + \beta\gamma)}{\gamma^2 + \delta^2} = \frac{\alpha\gamma + \beta\delta}{\gamma^2 + \delta^2} + i\frac{\beta\gamma - \alpha\delta}{\gamma^2 + \delta^2}$$

La expresión anterior nos proporciona la forma de dividir números complejos: se trata de multiplicar numerador y denominador por el conjugado del denominador y realizar las operaciones correspondientes.

**Ejemplo 1.2**

$$\frac{2 + 3i}{5 - 2i} = \frac{(2 + 3i)(5 + 2i)}{(5 - 2i)(5 + 2i)} = \frac{4 + 19i}{29} = \frac{4}{29} + \frac{19}{29}i$$

$$\frac{1 + i}{1 + 2i} = \frac{(1 + i)(1 - 2i)}{(1 + 2i)(1 - 2i)} = \frac{3 - i}{5} = \frac{3}{5} - \frac{1}{5}i$$

$$\frac{i}{2 - i} = \frac{i(2 + i)}{(2 - i)(2 + i)} = \frac{-1 + 2i}{5} = -\frac{1}{5} + \frac{2}{5}i$$

**Nota 1.2**

Las operaciones suma y producto de números complejos satisfacen las propiedades habituales de asociatividad, conmutatividad, existencia de elemento neutro, elemento simétrico (el opuesto en el caso de la suma y el inverso en el caso del producto) y propiedad distributiva del producto respecto de la suma (véase el apéndice A). Como consecuencia, $\mathbb{C}$, al igual que $\mathbb{R}$ ó $\mathbb{Q}$, tiene estructura de **cuerpo conmutativo**. Sin embargo es importante notar que, si bien $\mathbb{R}$ y $\mathbb{Q}$ son cuerpos **ordenados** (con la relación de orden $\leq$), $\mathbb{C}$ no lo es, es decir, no tiene sentido la expresión $z_1 \leq z_2$ para $z_1, z_2 \in \mathbb{C}$ (salvo que sean números reales).

1 2

## REPRESENTACIÓN GRÁFICA: MÓDULO Y ARGUMENTO

Un número complejo $z = \alpha + i\beta$ puede considerarse como un par ordenado de números reales $(\alpha, \beta)$. Estamos acostumbrados a representar geométricamente un par de este tipo como puntos en el plano cartesiano, siendo $\alpha$ la abscisa y $\beta$ la ordenada. Así pues, la identificación entre el número complejo $z$ y el par $(\mathrm{Re}(z), \mathrm{Im}(z))$ nos permite representar gráficamente un número complejo como un punto del plano, que denominaremos **plano complejo**,[4] en el que el eje de abscisas pasa a llamarse **eje real** y el eje de ordenadas **eje imaginario** (ver figura 1.1a).

Alternativamente, podemos representar también los números complejos como vectores en el plano, de manera que el número complejo $z = \alpha + i\beta$ puede identificarse con el vector que une los puntos $(0, 0)$ con $(\alpha, \beta)$ en el plano (ver figura 1.1b). Esta otra representación permite identificar elementos importantes de un número complejo como son su **módulo** y su **argumento**.

**Definición 1.4**

Dado un número complejo $z = \alpha + i\beta$ se define su *módulo*, que se notará como $|z|$, por

$$|z| = \sqrt{z\bar{z}} = \sqrt{\alpha^2 + \beta^2}$$

Como se puede observar, el módulo de un número complejo coincide con el módulo del vector (la longitud) que lo representa (ver figura 1.1b), y es igual a

[4] También conocido como *plano de Argand*, aunque en realidad fue el noruego Caspar Wessel en 1796 el primero que mostraba esta representación gráfica de los números complejos.

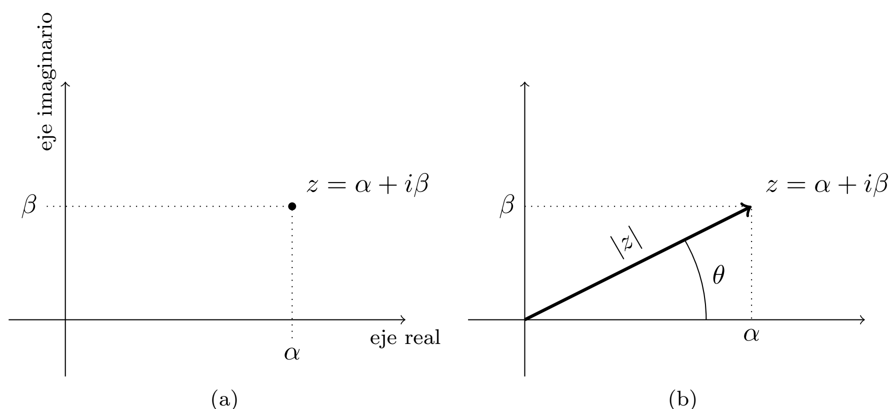



Figura 1.1: Representación gráfica de un número complejo

la distancia entre el punto del plano correspondiente al número complejo y el origen, siendo por tanto un número real positivo salvo en el origen, cuyo módulo, obviamente, es cero.

**Nota 1.3**

Si $z \in \mathbb{R}$, entonces su módulo coincide con su valor absoluto, de ahí que la notación empleada sea la misma.

Las principales propiedades del módulo se resumen en el siguiente resultado:

**Proposición 1.2**

(i) $|z| = |\overline{z}| = |-z|$, $\forall z \in \mathbb{C}$.

(ii) $|z_1 z_2| = |z_1|\,|z_2|$, $\forall z_1, z_2 \in \mathbb{C}$.

(iii) $\operatorname{Re}(z) \leq |\operatorname{Re}(z)| \leq |z|$.

(iv) $\operatorname{Im}(z) \leq |\operatorname{Im}(z)| \leq |z|$.

(v) **Desigualdad triangular**: $|z_1 + z_2| \leq |z_1| + |z_2|$, $\forall z_1, z_2 \in \mathbb{C}$.

(vi) $|z_1 - z_2| \geq ||z_1| - |z_2||$, $\forall z_1, z_2 \in \mathbb{C}$.

Para la demostración véanse los ejercicios 20 y 21.

Atendiendo a la figura 1.1b, para cada número complejo $z \in \mathbb{C}\backslash\{0\}$ también podemos considerar el ángulo $\theta$ que forma el vector que define $z$ con el semieje real positivo. Este ángulo se toma en radianes y se usa el criterio habitual de signos, que considera ángulos positivos los recorridos en sentido antihorario y ángulos negativos los del sentido horario. Entonces, la conocidas relaciones trigonométricas conducen a

$$\operatorname{Re}(z) = |z| \cos\theta, \quad \operatorname{Im}(z) = |z| \operatorname{sen}\theta \tag{1.1}$$

Debido a la periodicidad de las funciones reales trigonométricas, la igualdad anterior sigue siendo válida para $\theta + 2k\pi$, para cada $k \in \mathbb{Z}$. Teniendo esto presente podemos dar la siguiente definición:

**Definición 1.5**

Dado un número complejo $z = \alpha + i\beta \neq 0$ se define el *argumento* de $z$, y se notará por $\arg(z)$, al conjunto

$$\arg(z) = \{\theta \in \mathbb{R} : \alpha = |z| \cos\theta,\ \beta = |z| \operatorname{sen}\theta\}$$

El argumento de 0 no está definido.

Se denomina *argumento principal* de $z \in \mathbb{C}\backslash\{0\}$, y se notará por $\operatorname{Arg}(z)$, al único número $\theta \in \arg(z)$ tal que $-\pi < \theta \leq \pi$.

Es importante resaltar el carácter no unívoco del argumento de un número complejo. Es decir, el argumento no es un único valor, sino un conjunto de ellos. Este hecho tiene importantes consecuencias en la definición de ciertas funciones complejas (que no serán tratadas en este texto) y en el cálculo de raíces que veremos en la sección 1.4.

**Nota 1.4**

Gracias a la representación gráfica, la operación de conjugación puede verse de forma sencilla en términos del módulo y el argumento. Más concretamente, el conjugado de un número complejo corresponde a un vector simétrico respecto del eje real, por lo tanto posee el mismo módulo y opuesto argumento, es decir,

$$|z| = |\overline{z}|, \quad \arg(z) = -\arg(\overline{z})$$

Del mismo modo, la suma de números complejos corresponde a la suma de vectores que el lector conocerá de cursos anteriores, la conocida como *regla del paralelogramo*.

**Ejemplo 1.3**

Representemos en el plano complejo los números 3, $\frac{\sqrt{3}}{2} - \frac{1}{2}i$ y $-2+2i$ (véase la figura 1.2).

El módulo de cada uno de ellos es $|3| = 3$, $|\frac{\sqrt{3}}{2} + \frac{1}{2}i| = 1$ y $|-2+2i| = \sqrt{8}$ y sus argumentos principales:

$$\operatorname{Arg}(3) = 0, \quad \operatorname{Arg}(\tfrac{\sqrt{3}}{2} + \tfrac{1}{2}i) = -\tfrac{\pi}{6}, \quad \operatorname{Arg}(-2+2i) = \tfrac{3\pi}{4}$$

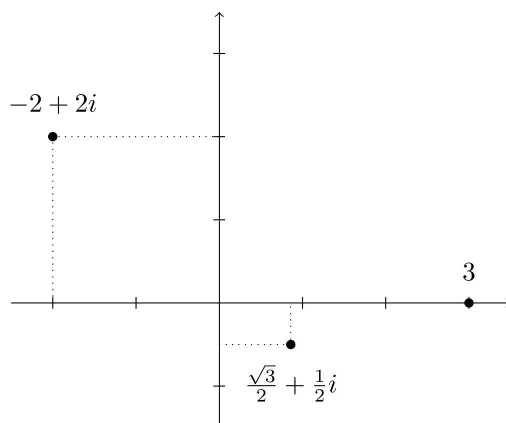



Figura 1.2: Representación de números complejos del ejemplo 1.3

## 1.3 FORMA TRIGONOMÉTRICA Y FORMA POLAR

Si $z \neq 0$ es un número complejo, para cada $\theta \in \arg(z)$, en virtud de (1.1), podemos escribir

$$z = |z|(\cos\theta + i \operatorname{sen}\theta) \tag{1.2}$$

De hecho, también es cierto si $z = 0$, para cualquier $\theta \in \mathbb{R}$.

**Definición 1.6**

A la expresión (1.2) se le denomina *forma trigonométrica* de un número complejo.

**Definición 1.7**

Dado $\theta \in \mathbb{R}$, se define $e^{i\theta} = \cos\theta + i \operatorname{sen}\theta$, que se conoce como *fórmula de Euler*.

Si ahora usamos (1.2) se tiene

$$z = |z|e^{i\theta}, \quad \theta \in \arg(z) \tag{1.3}$$

**Definición 1.8**

A la expresión (1.3) se le denomina *forma polar*[5] de un número complejo.

La forma polar de los números complejos facilita la multiplicación y división de los mismos gracias a las propiedades de la exponencial. Así, si $z_1 = r_1 e^{i\theta_1}$ y $z_2 = r_2 e^{i\theta_2}$ entonces

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}, \qquad \frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}$$

Esto nos permite representar gráficamente el producto y el cociente de números complejos. Así, el producto de dos números complejos tiene como módulo, el producto de los módulos, y como argumento, la suma de los argumentos. Es decir, si $z, w \in \mathbb{C}\backslash\{0\}$ entonces

$$\arg(zw) = \arg(z) + \arg(w)$$

No obstante es importante resaltar que la igualdad anterior es una igualdad entre conjuntos.[6] Es decir, se verifica que

$$\arg(z^2) = \arg(z) + \arg(z)$$

pero este hecho no es cierto para los argumentos principales. Por ejemplo, $\operatorname{Arg}((-i)^2) = \operatorname{Arg}(-1) = \pi$ pero $\operatorname{Arg}(-i) + \operatorname{Arg}(-i) = -\pi$.

[5] En ocasiones, para un número complejo escrito en forma polar como $z = re^{i\theta}$ se usa la notación $z = r_\theta$, en la que es habitual expresar el argumento en grados.

[6] Dados dos conjuntos $A$ y $B$, se define la suma $A + B$ como el conjunto $A + B = \{a + b : a \in A,\ b \in B\}$

1 4

## POTENCIA Y RAÍZ N-ÉSIMA DE UN NÚMERO COMPLEJO

La potencia natural de un número complejo se calcula fácilmente usando la forma polar; si $n \in \mathbb{N}$,

$$z = re^{i\theta} \Rightarrow z^n = r^n e^{in\theta}$$

Usando la forma trigonométrica se puede deducir la conocida como *fórmula de de Moivre*[7]

$$\cos(n\theta) + i \operatorname{sen}(n\theta) = (\cos\theta + i \operatorname{sen}\theta)^n \tag{1.4}$$

Por otro lado, dados $w \in \mathbb{C}$ y $n \in \mathbb{N}$, se define la *raíz n-ésima* de $w$ como el número complejo $z$ tal que $z^n = w$. Para realizar el cálculo hemos de usar nuevamente la forma polar, pero en este caso **es necesario usar todos los valores del argumento**. Más concretamente, si $w = re^{i\theta}$, escribiremos $w = re^{i(\theta+2k\pi)}$, de modo que

$$z^n = w \Rightarrow z = \sqrt[n]{r} e^{i\frac{\theta+2k\pi}{n}} \tag{1.5}$$

El resultado tiene como módulo $\sqrt[n]{r}$, que es la raíz n-ésima de un número real positivo, mientras que para cada valor de $k \in \mathbb{Z}$ aparecen, inicialmente, infinitos argumentos. Sin embargo, atendiendo a los ángulos que corresponden a estos argumentos observamos que, si tomamos cualesquiera $n$ valores consecutivos de $k$, obtenemos $n$ ángulos distintos, mientras que para el resto, los ángulos obtenidos vuelven a repetirse. Es decir, la raíz n-ésima de un número complejo da lugar a $n$ valores distintos, los cuales pueden obtenerse con $n$ valores consecutivos de $k$; típicamente tomaremos $k = 0, 1, \dots, n-1$.

Es más, atendiendo a los argumentos obtenidos se observa que las raíces n-ésimas forman un polígono regular de $n$ lados centrado en el origen y de radio $\sqrt[n]{r}$ (véase la figura 1.3).

**Ejemplo 1.4**

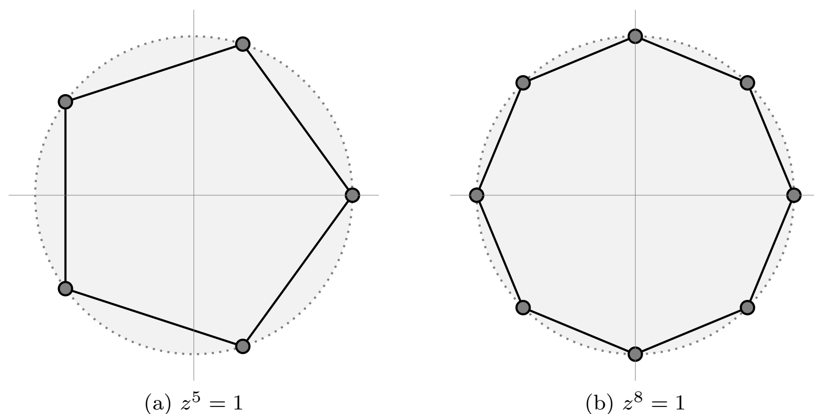
Calculemos los números complejos tales que $z^8 = 1$ usando (1.5). En primer lugar obtenemos la forma polar de $1 = e^{i2k\pi}$, usando todos los argumentos. Así

$$z^8 = 1 = e^{i2k\pi} \Rightarrow z = e^{i\frac{2k\pi}{8}},\ k = 0, \dots, 7$$

obteniéndose las raíces:

$$e^{i0} = 1, \quad e^{i\frac{\pi}{4}} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, \quad e^{i\frac{\pi}{2}} = i, \quad e^{i\frac{3\pi}{4}} = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i,$$
$$e^{i\pi} = -1, \quad e^{\frac{5\pi}{4}} = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i, \quad e^{\frac{3\pi}{2}} = -i, \quad e^{i\frac{7\pi}{4}} = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i.$$

[7] Denominada así por el matemático francés Abraham de Moivre. La fórmula aparece publicada en un artículo suyo en 1722.



(a) $z^5 = 1$ (b) $z^8 = 1$

Figura 1.3: Raíces de la unidad

Las raíces aparecen representadas en la figura 1.3b.

**Ejemplo 1.5**

¿Cuáles de los siguientes números complejos $1+i$, $-1+i$, $-1-i$ y $1-i$ son raíces décimas de $32i$?

Si escribimos cada uno de estos números en forma polar,

$$1+i = \sqrt{2}e^{i\frac{\pi}{4}}, \quad -1+i = \sqrt{2}e^{i\frac{3\pi}{4}}, \quad -1-i = \sqrt{2}e^{-i\frac{3\pi}{4}}, \quad 1-i = \sqrt{2}e^{-i\frac{\pi}{4}}$$

calcular su potencia décima es fácil:

$$(1+i)^{10} = 32e^{i\frac{5\pi}{2}} = 32i, \quad (-1+i)^{10} = 32e^{i\frac{15\pi}{2}} = -32i,$$
$$(-1-i)^{10} = 32e^{-i\frac{15\pi}{2}} = 32i, \quad (1-i)^{10} = 32e^{-i\frac{5\pi}{4}} = -32i.$$

Luego los números buscados son $1+i$ y $-1-i$. Nótese que al tratarse del cálculo de potencias naturales no es necesario tener en cuenta todos los argumentos a diferencia de lo que ocurre cuando calculamos raíces.

Para finalizar, daremos un resultado que será de vital importancia en el tema 6 y que afirma que $\mathbb{C}$ es un cuerpo algebraicamente cerrado, es decir, que

todo polinomio en $\mathbb{C}$ tiene al menos una raíz. Por ejemplo, es fácil ver que $\mathbb{R}$ no es algebraicamente cerrado, pues el polinomio $x^2+1$ no tiene raíces reales. Este resultado, conocido como *Teorema fundamental del Álgebra*, afirma lo siguiente:

**Teorema 1.1 (Teorema fundamental del Álgebra)**

Si $p(x) = a_0 + a_1x + \cdots + a_nx^n$ es un polinomio con coeficientes complejos (esto es, $a_i \in \mathbb{C}$, $\forall i$), entonces la ecuación $p(x) = 0$ tiene $n$ raíces contando sus multiplicidades.

Aunque existe una fácil demostración de este resultado, la maquinaria empleada para la misma excede los objetivos de este curso.[8]

El interés principal que para nosotros tiene este resultado ocurre cuando el polinomio tiene coeficientes reales (es decir, $a_i \in \mathbb{R}$, $\forall i$). En tal caso, el polinomio tendrá $n$ raíces (reales o complejas), y además, si $z$ es una raíz compleja de $p$, entonces $\overline{z}$ también es raíz de $p$ (véase el ejercicio 24).

## 1.5 CÁLCULO CON PYTHON

En esta sección mostraremos cómo usar Python para operar con números complejos, a la vez que profundizaremos en el empleo del módulo SymPy presentado en el apéndice B. Es esencial que el lector posea un mínimo de conocimientos del lenguaje (que exponemos en el citado apéndice), por lo que remitimos a su lectura antes de abordar cualquiera de las secciones dedicadas a Python.

Por otra parte, es muy conveniente que el lector realice los ejemplos que exponemos aquí para habituarse al intérprete Python. Para facilitar la comprensión de los ejemplos hemos numerado de forma consecutiva las líneas de órdenes introducidas en el intérprete, simulando una determinada sesión interactiva, por lo que de un ejemplo al siguiente se conservan las importaciones de los módulos realizadas previamente. Cuando los números de línea comienzan de nuevo, significa que hemos abierto una nueva sesión, y por tanto debemos volver a cargar ciertas funciones.

[8]El primer matemático en enunciar un resultado parecido (que todas las ecuaciones de grado $n$ tienen $n$ raíces) fue el francés Albert Girard en 1629, aunque no menciona que tales soluciones puedan ser complejas. El resultado se aceptó como un hecho evidente por la comunidad matemática. El primer intento serio de demostración se debió al también francés Jean Le Rond d'Alambert en 1746, pero la prueba tiene algunos puntos negros. Aunque es aceptado que la primera demostración del resultado es debida a Gauss en 1799 en su tesis doctoral, la prueba no es completamente rigurosa según los estándares actuales. Gauss dio otras tres pruebas del resultado a lo largo de su vida; la última apareció en 1849, en el último artículo que escribió, cincuenta años después de su primera demostración.

### 1.5.1 Operaciones básicas

Comenzaremos con ejemplos de operaciones algebraicas sencillas con Python. Como se ve en la sección B.2, los números complejos están accesibles en Python y operar con ellos es fácil. Realicemos las siguientes operaciones:

$$(2-i)-(6+2i) = -4-3i$$

$$(3+2i)(1-5i) = 13-13$$

$$\frac{1+i}{1+2i} = \frac{3}{5} - \frac{1}{5}i$$

```
>>> 2-1j-(6+2j)
(-4-3j)
>>> (3+2j)*(1-5j)
(13-13j)
>>> (1+1j)/(1+2j)
(0.59999999999999998-0.20000000000000001j)
```

Como ocurre con las calculadoras, en el momento en el que aparecen números racionales el cálculo se vuelve engorroso. Afortunadamente tenemos el módulo SymPy para operar de forma exacta. En este módulo tenemos definida la unidad imaginaria como `I`, pero hemos de usar el operador de multiplicación (a diferencia del uso con `j`): por ejemplo,

$$\frac{1+i}{1+2i} = \frac{3}{5} - \frac{1}{5}i$$

```
>>> from sympy import I,simplify
>>> (1+I)/(1+2*I)
(1 + I)/(1 + 2*I)
>>> simplify(_)
3/5 - I/5
```

Como podemos observar, la operación $\frac{1+i}{1+2i}$ no se realiza al momento (es tratada como una expresión simbólica), por lo que debemos simplificarla. Nótese el uso del guión bajo (o *underscore*) para referirnos al último cálculo realizado. Nótese también que solo hemos importado las funciones del módulo que necesitamos. Podríamos haber escrito directamente

```
>>> simplify(6/(2-I))
12/5 + 6*I/5
```

que equivale a

$$\frac{6}{2-i} = \frac{12}{5} + \frac{6}{5}i$$

El módulo SymPy también dispone de funciones para calcular el conjugado, el módulo y el argumento de un número complejo:

```
>>> from sympy import arg,abs,conjugate,sqrt,I
>>> a=sqrt(3)+I
>>> conjugate(a)
3**(1/2) - I
>>> abs(a)
2
>>> arg(a)
pi/6
```

que realiza los cálculos

$$\overline{\sqrt{3}+i} = \sqrt{3} - i \qquad \left|\sqrt{3}+i\right| = 2 \qquad \arg(\sqrt{3}+i) = \frac{\pi}{6}$$

Observemos que hemos importado del módulo SymPy, además de las funciones que vamos a usar, también la función `sqrt` (raíz cuadrada). Recordamos aquí que esta función no viene por defecto con el núcleo de Python y es necesario importarla. ¿Qué hubiera ocurrido si, en lugar de usar la función `sqrt` de SymPy, usamos la del módulo `math`:

```
>>> from sympy import conjugate,I
>>> from math import sqrt
>>> a=sqrt(2)+I
>>> conjugate(a)
1.73205080756888 - I
```

Insistimos en la ventaja de hacer las operaciones con funciones de SymPy, que, en la medida de lo posible, realizará las operaciones de forma exacta. Como hemos visto en el ejemplo anterior, SymPy es capaz de reconocer ciertos ángulos habituales, pero obviamente, no hace milagros:

```
>>> from sympy import arg,abs,sqrt,I
>>> a=sqrt(2)+I
>>> abs(a)
3**(1/2)
>>> arg(a)
atan(2**(1/2)/2)
```

que interpretamos como

$$\left|\sqrt{2}+i\right| = \sqrt{3} \qquad \arg(\sqrt{2}+i) = \arctan\frac{\sqrt{2}}{2}$$

Si necesitamos el valor real de alguna de las operaciones efectuadas con SymPy, podemos usar el método `evalf`

```
>>> arg(a).evalf()
0.615479708670387
```

esto es,

$$\arctan \frac{\sqrt{2}}{2} = 0.615479708670387$$

que al ser un método y no una función, no precisa importación.

### 1.5.2 Operaciones avanzadas

El módulo SymPy permite realizar cálculos más sofisticados con los números complejos: por ejemplo, reconoce la fórmula de Euler, y es capaz de simplificar operaciones trigonométricas:

```
>>> from sympy import Symbol,E,I,re
>>> x=Symbol('x',real=True)
>>> a=E**(I*x)
>>> b=a.expand(complex=True)
>>> b
I*sin(x) + cos(x)
>>> c=a**3
>>> d=c.expand(complex=True).expand(trig=True)
>>> d
-3*cos(x) - I*sin(x) + 4*I*cos(x)**2*sin(x)
+ 4*cos(x)**3
>>> re(d)
-3*cos(x) + 4*cos(x)**3
>>> f=c.expand(complex=True)
>>> re(f)
cos(3*x)
>>> (re(f)-re(d)).expand(trig=True)
0
```

Nótese el uso de la etiqueta adicional `real=True` en la función `Symbol`, cuyo significado es evidente: asume que la variable simbólica `x` es real. El número $e$ viene dado por `E` y la parte real es `re`. El método `expand` sirve para desarrollar expresiones, y admite etiquetas adicionales para usar desarrollos con complejos o trigonométricos (admite otras varias como logaritmos, potencias, etc.).

Los cálculos prueban la fórmula:

$$\cos(3x) = -3\cos(x) + 4\cos^3(x)$$

¿Puede el lector encontrar una fórmula análoga para $\operatorname{sen}(3x)$?

El módulo SymPy también permite encontrar las raíces complejas de polinomios. La función `solve` nos proporciona las raíces de una ecuación del siguiente modo:

```
>>> from sympy import Symbol,solve
>>> x=Symbol('x')
>>> solve(x**4-1,x)
[1, -1, -I, I]
```

esto es,

$$x^4 - 1 = 0 \Rightarrow x = 1, -1, i, -i$$

Si el resultado esperado es amplio podemos guardarlo en una lista

```
>>> from sympy import I
>>> a=solve(x**4-I,x)
>>> a
[-sin(pi/8) + I*cos(pi/8),
 -I*cos(pi/8) + sin(pi/8),
 -cos(pi/8) - I*sin(pi/8),
 I*sin(pi/8) + cos(pi/8)]
```

y luego podemos usar un bucle para, por ejemplo, obtener los valores numéricos de cada una de las soluciones:

```
>>> for b in a:
...     b.evalf()
...
-0.38268343236509 + 0.923879532511287*I
0.38268343236509 - 0.923879532511287*I
-0.923879532511287 - 0.38268343236509*I
0.923879532511287 + 0.38268343236509*I
```

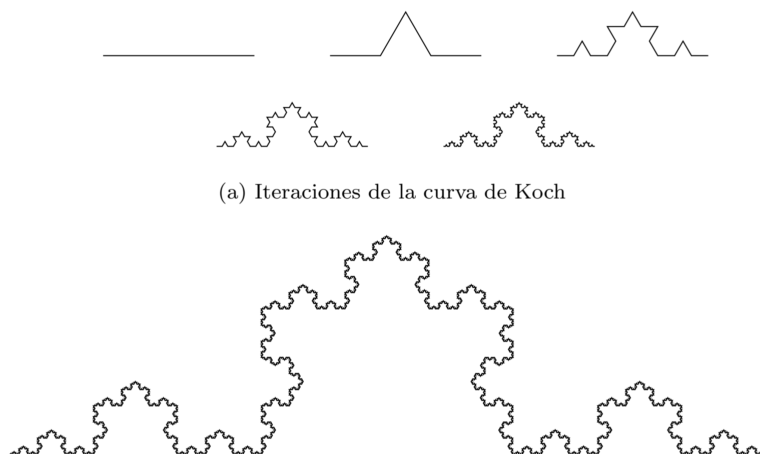
## 1 6
## UNA BREVE INCURSIÓN EN EL MUNDO FRACTAL

Los fractales[9] son objetos geométricos cuya estructura, generalmente muy irregular, se repite en diferentes escalas. Esencialmente podríamos decir que un fractal es un objeto de *dimensión fraccionaria*, aunque sería necesario dar una definición concreta del concepto de dimensión al que nos referimos.

Los primeros ejemplos de fractales ya aparecen a finales del siglo XIX, con la función de Weierstrass descubierta en 1872, que es una función continua en todos sus puntos pero que no es derivable en ninguno. El grafo de esta función es un objeto fractal. Posteriormente aparecieron otros objetos, como la *curva de Koch*[10] (véase la figura 1.4), que se trata de una curva continua que no posee tangentes (es decir, que no es derivable) obtenida mediante un procedimiento recursivo bastante simple: dado un segmento, se divide en tres partes, y la parte central se sustituye por dos segmentos de igual longitud a los anteriores formando un triángulo sin la base (véase la figura 1.4a). El proceso se repite para cada uno de los segmentos de la nueva figura, y el objeto obtenido en el límite es la mencionada curva.

[9] El término fue acuñado por el matemático francés de origen polaco Benoît Mandelbrot, en un artículo publicado en 1967 por la revista *Science* titulado *¿Cuánto mide la costa de Gran Bretaña?*

[10] Descubierta por el matemático sueco Helge von Koch en 1904.



(a) Iteraciones de la curva de Koch

Figura 1.4: Curva de Koch

Es fácil darse cuenta de que la longitud de esta curva es infinita, pues en cada iteración estamos incrementando su longitud: por ejemplo, si el segmento inicial (el de arriba a la izquierda en la figura 1.4a) tiene longitud 1, la primera iteración tendrá longitud $\frac{4}{3}$, la segunda $\frac{16}{9}$, la tercera $\left(\frac{4}{3}\right)^3$, ..., y en la $n$-ésima iteración, la longitud será de $\left(\frac{4}{3}\right)^n$, de modo que la longitud límite será infinita. Este hecho nos invita a pensar que, de algún modo, este objeto no es unidimensional. Usando la definición oportuna se puede encontrar la dimensión fractal de este objeto que es $\frac{\ln 4}{\ln 3} = 1.261\ldots$

Este tipo de objetos fueron considerados durante cierto tiempo por la comunidad matemática como anomalías artificiales[11] y apenas recibieron atención. Tras el nacimiento de la *Teoría del Caos*,[12] el interés por estos objetos se renueva, ya que los fractales aparecen vinculados a la evolución de sistemas complejos. Posteriormente, y en especial a partir de la publicación del libro de Mandelbrot, *La Geometría Fractal de la Naturaleza* (1982), los fractales son usados para describir la complejidad de ciertas formas en la naturaleza, desde las nubes hasta las redes neuronales del cerebro.

¿Qué relación hay entre los números complejos y los fractales? Uno de los fractales más conocidos, el denominado conjunto de Mandelbrot,[13] es un objeto del plano complejo que se genera de forma sencilla. Se considera un número

[11] Una galería de "monstruos", como los denominó el matemático francés Henri Poincaré.

[12] Surgida a partir del estudio de un modelo climático por el matemático estadounidense Edward Lorenz en 1963.

[13] Las primeras imágenes del mismo se obtuvieron en 1978 por Robert Brook y Peter Matelski, pero fueron el matemático francés Adrien Douady y su discípulo John H. Hubbard quienes demostraron muchas propiedades fundamentales y lo nombraron así en honor a Mandelbrot.

complejo cualquiera $c$, y se construye una sucesión por recurrencia del siguiente modo:

$$z_0 = 0, \qquad z_{n+1} = z_n^2 + c, \quad n \geq 0$$

En función del número $c$ escogido, esta sucesión puede o no estar acotada. Por ejemplo, si $c = 2$, la sucesión que se genera es $0, 2, 6, 38, \dots$ que es claramente no acotada; pero si $c = -1$, la sucesión queda $0, -1, 0, -1, \dots$ que sí está acotada. El conjunto de Mandelbrot está formado por los números complejos $c$ tales que la sucesión anterior permanece acotada. Es decir, $-1$ es un punto que está en el conjunto de Mandelbrot, pero 2 no pertenece a dicho conjunto. Si analizamos el comportamiento de la sucesión con todos los puntos del plano complejo obtenemos la figura 1.5a, en la que se aprecia el aspecto general del conjunto de Mandelbrot. Las figuras 1.5b y 1.5c muestran aproximaciones (las indicadas en los recuadros) con más detalle. Podemos ver la "autosimilitud" característica de los fractales (se repiten en escalas cada vez más pequeñas) y el "orden dentro del caos" que subyace en los mismos.[14]

### 1 6 1 Generación del conjunto de Mandelbrot con Python

Adjuntamos aquí un breve programa en Python con el que obtener los gráficos aquí mostrados. Para su funcionamiento se precisa del módulo `matplotlib`[15] para poder generar gráficos con Python.

```
#! /usr/bin/python

from numpy import array,zeros
from matplotlib.pyplot import imshow,xticks,yticks,show

def mandelplot(size, limit, xint, yint):

        img = zeros([size, size], int)

        xamp=xint[1]-xint[0]
        yamp=yint[1]-yint[0]

        for y in range(size):
                for x in range(size):
                        c = complex(x/float(size)*xamp +
                            xint[0],y/float(size)*yamp + yint
                            [0])
                        z = c
                        for i in range(limit):
                                z = z**2 + c
```

[14]Precisamente esta característica de los fractales, la aparición de cierta regularidad dentro del caos, fue lo que llamó la atención de Mandelbrot cuando intentaba descifrar la causa de la existencia de determinado ruido en las líneas telefónicas que usaban para transmitir información en la red de ordenadores de la compañía IBM, en la cual trabajaba.

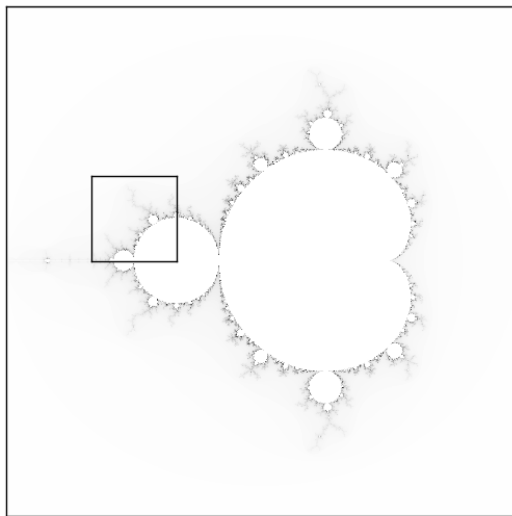
[15]Se puede descargar desde `http://matplotlib.sourceforge.net/`.

```
19                                  img[y,x] += 1
20                                  if abs(z) > 2:
21                                          break
22                          else:
23                                  img[y,x] = 0
24
25          img = array(img/float(img.max()))
26
27          asp=yamp/xamp
28
29          imshow(img,interpolation='bilinear',origin='lower',
                cmap='binary',aspect=asp)
30
31          xticks([])
32          yticks([])
33          show()
34
35          return
36
37
38  puntos = 1000
39  limite = 250
40  xint=[-2.,1.]
41  yint=[-1.5,1.5]
42
43  mandelplot(puntos,limite,xint,yint)
```

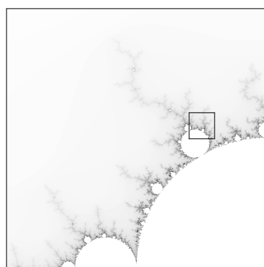
El funcionamiento del programa consiste en, dado un determinado conjunto de puntos del plano complejo, analizar las iteraciones de la sucesión que define al conjunto en cada uno de esos puntos. Si en algún momento de la iteración el módulo es mayor que 2, automáticamente sabemos que ese número complejo no va a pertenecer al conjunto de Mandelbrot. Por el contrario, si después de haber realizado un cierto número de iteraciones, el módulo sigue siendo menor que 2, consideramos que ese punto está dentro del conjunto.

Analicemos con más detenimiento el código: La línea 1 es el *shebang* comentado en la sección B.1.1. Las líneas 3 y 4 cargan las funciones a usar desde los módulos y entre las líneas 6 y 35 definimos la función que hace todo el trabajo y que invocamos en la línea 43.

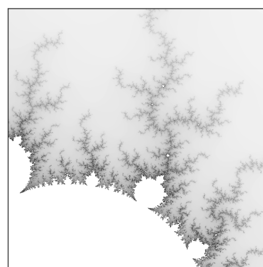
Los argumentos de entrada de la función son: `size`, `limit`, `xint` e `yint` que son definidos en las líneas 38 a 41. Estos dos últimos son dos listas que se refieren al dominio en el que se van a realizar los cálculos: `xint` nos da la cota inferior y superior para la parte real de los números complejos a considerar, e `yint` define las cotas para la parte imaginaria. El parámetro `size` define el número de subdivisiones que vamos a realizar en cada intervalo `xint` e `yint` y el parámetro `limit` se refiere al número máximo de veces que vamos a realizar la iteración que define al conjunto.



(a) Rectángulo $[-2, 1] \times [-\frac{3}{2}, \frac{3}{2}]$



(b) Rectángulo $[-\frac{3}{2}, -1] \times [0, \frac{1}{2}]$



(c)

Figura 1.5: El conjunto de Mandelbrot

La función, una vez introducidos los parámetros, crea una matriz de ceros de tamaño `size`×`size` (línea 8); en las líneas 10 y 11 calcula la amplitud de la región a dibujar y entre las líneas 13 y 23 está el doble bucle que recorre cada uno de los puntos `c` de la región a considerar (definidos en la línea 15).

Para cada punto, se entra en un bucle `for-else` (líneas 17–23) en el que se analiza si el punto está o no en el conjunto de Mandelbrot. Para ello se van construyendo los elementos de la sucesión (línea 18) hasta la iteración `limit` (como máximo). Si en algún elemento de la sucesión el módulo es mayor que 2 entonces se interrumpe el bucle (líneas 20–21) y se pasa a analizar el siguiente punto. Nótese que el elemento correspondiente a la matriz `img` va sumando 1 cada vez que se construye un elemento de la sucesión (línea 19). Así, si interrumpimos el bucle en un punto porque el módulo es mayor que 2, habremos asignado a la matriz el índice de la sucesión en el que se ha obtenido la condición de salida.

Por el contrario, si el bucle `for` finaliza sin un `break` (es decir, los elementos de la sucesión se mantienen todos con módulo menor o igual que 2), entonces se ejecuta el bloque `else` (nótese el sangrado), poniendo el elemento de la matriz a 0 (línea 23).

Al finalizar el doble bucle (líneas 17–23), el programa habrá construido una matriz en la que cada elemento guarda un índice entero para cada punto del rectángulo del plano complejo analizado. Este índice vale 0 si el punto está en el conjunto, y es distinto de 0 si no lo está. Índices altos indican que ha costado decidir si el punto está o no en el conjunto, mientras que índices bajos corresponden a puntos que rápidamente han sido descartados por su pertenencia al conjunto. Este rango de índices nos va a permitir dar color al dibujo. Si solo hubiéramos indicado la pertenencia al conjunto con un 1 o un 0 apreciaríamos menos detalles en la gráfica. El lector puede probar con distintos valores del parámetro `limit` para corroborar esto.

La línea 25 normaliza los valores de la matriz `img` dividiendo por el mayor valor de ésta (lo que se obtiene con `img.max()`). Obsérvese también la necesidad de convertir a un número real con `float`. La matriz ahora tendrá valores reales entre 0 y 1.

Finalmente, la función `imshow` de `matplotlib` crea un dibujo a partir de la matriz coloreando cada píxel en función del valor (el coloreado usado en este caso ha sido *binary*). Las funciones `xtick` e `ytick` eliminan la información sobre los ejes y `show()` muestra el resultado.

El código de la función (líneas 6–35) no se ejecuta hasta que la función no es llamada en la línea 43. Las líneas 38–41 definen los parámetros con los que se ha obtenido la figura 1.5a. El lector puede probar modificando estos valores para obtener distintas partes del conjunto, con diferentes niveles de aproximación.

1 7

## EJERCICIOS

### Ejercicios de repaso

**E.1** Realizar las siguientes operaciones con números complejos:

$$(3-2i)+(1-5i) \qquad (2-2i)+3i \qquad (1+2i)-(-1+2i)$$
$$(2+5i)(3+i) \qquad (1+i)(1-i) \qquad 2i-(5+5i)(3-6i)$$
$$\frac{1+3i}{2-5i} \qquad \frac{4-7i}{3+5i} \qquad \frac{i}{2-i}$$
$$\frac{1}{i}-2i(2-i) \qquad \frac{i}{5}-\frac{2-i}{2+i} \qquad \frac{3i-1}{i}+1$$

**E.2** Obtener el conjugado de los siguientes números complejos:

$$i, \quad 3-2i, \quad \pi+e-i, \quad e-\pi, \quad \frac{1}{i}, \quad \tfrac{3}{2}i-1$$

**E.3** Calcular el módulo y el argumento principal de los siguientes números complejos y representarlos en el plano complejo:

$$-2, \quad -i, \quad 5-5i, \quad -3\sqrt{3}+3i, \quad -1-\sqrt{3}i, \quad -2-2i$$

**E.4** Escribir en forma polar y trigonométrica los números complejos del ejercicio 3.

**E.5** Escribir en forma cartesiana los siguientes números y representarlos en el plano complejo:

$$e^{i7\pi}, \quad 3e^{i\frac{\pi}{4}}, \quad 2e^{-i\frac{3\pi}{4}}, \quad \tfrac{1}{2}e^{i\frac{\pi}{6}}, \quad \tfrac{2}{3}e^{i\frac{5\pi}{4}}, \quad e^{i\frac{2\pi}{3}}$$

**E.6** Calcular las siguientes potencias:

$$i^5, \quad (1-\sqrt{3}i)^5, \quad (-2+2i)^8, \quad (\tfrac{1}{2}-\tfrac{\sqrt{3}}{2}i)^{10}$$

**E.7** Obtener en forma binomial todas las soluciones de las siguientes ecuaciones:

$$z^4=2, \quad z^3=\tfrac{1}{8}i, \quad z^6=-1$$

**E.8** Resolver las siguientes ecuaciones en $\mathbb{C}$:

$$z^2+5=0; \qquad z^2+z+1=0; \qquad z^3+z^2+z=0;$$
$$z^2+2z+5=0; \quad z^3-3z^2+5z=15; \quad z^3+2z^2+4z=-3.$$

### Problemas

**E.9** Escribir en forma cartesiana los números

$$\frac{(1+\sqrt{3}i)^3}{(\sqrt{3}+i)^2}, \quad i^{501}+i^{600}, \quad \sum_{k=0}^{100} i^k$$

Indicación: para el último usar la fórmula $\displaystyle\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}$

**E.10** ¿En qué cuadrante se encuentran los siguientes números?

$$(2000e^{\frac{3\pi}{4}i}+1)^7, \quad (5000e^{\frac{9\pi}{7}i}+2)^{11}$$

**E.11** Sea $z \in \mathbb{C}$ tal que $z^6 = -64$, $\mathrm{Re}(z) < 0$ y $\mathrm{Im}(z) < 0$. Calcular $z^4$.

**E.12** Calcular $w = (1-i)^{12} z^{-1}$, donde $z$ es tal que $z^4 + 1 = 0$, $\mathrm{Re}(z) > 0$ y $\mathrm{Im}(z) > 0$.

**E.13** Si $x \in \mathbb{R}$ y $n \in \mathbb{N}$, calcula el valor de

$$\left(\frac{1+\cos x + i \operatorname{sen} x}{1+\cos x - i \operatorname{sen} x}\right)^n$$

**E.14** Determinar los números complejos que satisfacen:

$$\text{(a) } z = |\overline{z}| \qquad \text{(b) } z = \overline{z}^2$$

**E.15** Halla las números complejos que forman los vértices de un hexágono regular de centro el origen, sabiendo que tiene uno de sus vértices en el número $i$.

**E.16** Describe los conjuntos de números complejos que satisfacen las ecuaciones:

(a) $|z| \leq 3$ (b) $|z-i| > 2$ $^\star$(c) $|z+1| + |z-1| = 4$

***E.17** Encontrar el lugar geométrico descrito por la ecuación $\left|\frac{z+1}{z-1}\right| = 2$.

### Ejercicios teóricos

**E.18** Probar que para cada $z \in \mathbb{C}$,

$$\mathrm{Re}(z) = \frac{z+\overline{z}}{2}, \quad \mathrm{Im}(z) = \frac{z-\overline{z}}{2i}$$

**E.19** Demostrar la Proposición 1.1.

**E.20** Probar (I)–(IV) de la Proposición 1.2.

***E.21** En este ejercicio se pretende demostrar la desigualdad triangular (véase (V) de la Proposición 1.2). Para ello:

(a) Probar que $2\alpha\beta \leq \alpha^2 + \beta^2$, $\forall \alpha, \beta \in \mathbb{R}$.

(b) Usar (a) para probar que

$$\alpha_1\alpha_2 + \beta_1\beta_2 \leq \sqrt{(\alpha_1^2+\beta_1^2)(\alpha_2^2+\beta_2^2)}, \quad \forall \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$$

(c) Usar (b) para probar que $z_1\overline{z_2} + \overline{z_1}z_2 \leq 2|z_1|\,|z_2|$, $\forall z_1, z_2 \in \mathbb{C}$.

(d) Deducir de (c) la desigualdad triangular y (VI) de la Proposición 1.2.

*E.22 Decidir si las siguientes afirmaciones son verdaderas o falsas:

(a) Los puntos $z$, $w$, $-z$ y $-w$ forman un cuadrado si y solo si $\frac{z}{w}$ es imaginario puro.

(b) Los puntos $z$, $w$ y 0 están alineados si y solo si $\frac{z}{w}$ es imaginario puro.

(c) Los puntos $z$, $w$ y 0 están alineados si y solo si $\frac{z}{w}$ es real.

*E.23 Sean $z_k$, $k = 0, \ldots, n-1$ las raíces n-ésimas de la unidad. Probar que

$$\sum_{k=0}^{n-1} \mathrm{Re}(z_k) = 0$$

Indicación: usar la ayuda del ejercicio 9

E.24 Probar que si $p(x) = a_0 + a_1x + \cdots + a_nx^n$ es un polinomio con coeficientes reales (esto es, $a_i \in \mathbb{R}$, $\forall i$) y $z_0$ es una raíz de $p$ entonces $\overline{z}_0$ también es raíz de $p$. ¿Es cierto esto si algún $a_i \in \mathbb{C}$?

## Ejercicios adicionales

E.25 Usa la fórmula de Euler en Python para obtener una identidad trigonométrica para $\mathrm{sen}(5x)$ y $\cos(5x)$.

E.26 Los **Conjuntos de Julia**[16] son fractales que se generan de forma similar al conjunto de Mandelbrot: se considera un número complejo $\xi$ fijo, y para cada $c \in \mathbb{C}$ se construye la sucesión

$$z_0 = c \qquad z_{n+1} = z_n^2 + \xi, \quad n \geq 0$$

El conjunto de Julia generado por $\xi$ es el conjunto de números complejos $c$ para los que la sucesión anterior permanece acotada. Adaptar el código en Python de la sección 1.6.1 para dibujar los conjuntos de Julia para $\xi = 0.72 - 0.196i$ y $\xi = -0.1 + 0.87i$.

[16] Llamados así en honor al matemático francés de origen argelino Gaston Julia.

# 2 Matrices y determinantes

Introducimos en este tema una de las herramientas fundamentales del Álgebra, las matrices, con las que trabajaremos a lo largo de todo el texto. Es probable que el lector haya tenido ya un contacto previo con estos objetos así como con los determinantes, por lo que este tema consistirá esencialmente en un repaso de tales contenidos aunque con un enfoque más abstracto. Básicamente recordaremos las operaciones entre matrices, el método de Gauss y el cálculo de determinantes, esperando que el lector adquiera una adecuada soltura operacional con ellos.

Puesto que el cálculo matricial es sin duda tedioso, principalmente cuando el tamaño de las matrices aumenta, incluimos también una sección dedicada a la realización de estas operaciones con Python. De este modo el lector tendrá la oportunidad de realizar los numerosos cálculos que involucren matrices a lo largo del texto con rapidez. No obstante, es importante resaltar la importancia de saber realizar tales operaciones sin la necesidad del ordenador. Por último, hacemos una pequeña incursión en una de las múltiples aplicaciones donde aparecen las matrices: la teoría de grafos.

## 2 1 MATRICES: PRIMERAS DEFINICIONES

En todos los temas que siguen trabajaremos con conjuntos numéricos con los que será necesario realizar cierto tipo de operaciones. Más concretamente, necesitaremos que tales conjuntos tengan estructura de cuerpo conmutativo (véase el apéndice A), que denotaremos genéricamente por $\mathbb{K}$, y que en la práctica puede ser $\mathbb{Q}$, $\mathbb{R}$ ó $\mathbb{C}$. En los casos en los que sea necesario precisar el cuerpo concreto con el que trabajar lo diremos expresamente.

A los elementos del cuerpo los denominaremos *escalares* y se denotarán habitualmente con letras griegas: $\alpha$, $\beta$, $\gamma$, etc. También usaremos la notación $\mathbb{K}^n$ para referirnos al producto cartesiano $\mathbb{K}\times \overset{n}{\cdots} \times\mathbb{K}$, es decir el conjunto de $n$-uplas $(\alpha_1, \dots, \alpha_n)$, donde cada $\alpha_i \in \mathbb{K}$.

**Definición 2.1**

Una ordenación rectangular

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

se denomina *matriz*[1] *de orden* $m \times n$ ($m$ *filas* y $n$ *columnas*), donde cada $a_{ij} \in \mathbb{K}$. Los elementos $a_{i1}, a_{i2}, \dots a_{in}$ forman la fila i-ésima, que denotaremos por $F_i$, mientras que los elementos $a_{1j}, a_{2j}, \dots, a_{mj}$ conforman la columna j-ésima, denotada por $C_j$. El conjunto de matrices de orden $m \times n$ con coeficientes en $\mathbb{K}$ se denota por $\mathcal{M}_{m\times n}(\mathbb{K})$. Para simplificar la escritura es frecuente notar a las matrices por $A = (a_{ij})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$ o simplemente $A = (a_{ij})$, donde el primer índice siempre denotará las filas y el segundo las columnas.

Una matriz $A \in \mathcal{M}_{1\times n}(\mathbb{K})$ se dirá matriz *fila* y si $A \in \mathcal{M}_{m\times 1}(\mathbb{K})$ se dirá matriz *columna*.

**Definición 2.2**

Dos matrices $A, B \in \mathcal{M}_{m\times n}(\mathbb{K})$ son *iguales* si los elementos correspondientes son iguales, es decir, $a_{ij} = b_{ij}$, $1 \leq i \leq m$, $1 \leq j \leq n$.

Las matrices aparecen en multitud de situaciones por tratarse de objetos eficaces para tratar información de una manera ordenada. El siguiente ejemplo es una muestra de ello.

**Ejemplo 2.1**

Una empresa fabrica pilas eléctricas en tres tamaños: A, B y C, y dos voltajes: $V_1$ y $V_2$. El número de piezas fabricadas cada día (en miles de unidades) viene

[1] El término fue acuñado por el matemático inglés James Joseph Sylvester en 1848, aunque ordenaciones de este tipo son conocidas desde la antigüedad; por ejemplo, hay constancia de la aparición de cuadrados mágicos en la literatura china hacia el 650 a.C.

dada por la matriz $F$, y el precio (en céntimos por unidad) en la matriz $P$:

$$F = \begin{matrix} & \begin{matrix} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{matrix} \\ \begin{matrix} V_1 \\ V_2 \end{matrix} & \begin{pmatrix} 20 & 19 & 18 \\ 22 & 19 & 21 \end{pmatrix} \end{matrix} \qquad P = \begin{matrix} & \begin{matrix} V_1 & V_2 \end{matrix} \\ \begin{matrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{matrix} & \begin{pmatrix} 70 & 120 \\ 45 & 65 \\ 60 & 50 \end{pmatrix} \end{matrix}$$

Como se puede observar, las matrices anteriores contienen la información sobre la producción diaria de cada tipo de pila y sus precios. A partir de estas matrices podemos extraer información implícitamente contenida en ellas. Por ejemplo, la suma de todos los elementos de la matriz $F$ nos proporciona la cantidad total de pilas fabricadas de todos los tipos. Si sumamos los elementos de cada columna obtendremos la cantidad de pilas fabricadas de los tipos A, B y C tanto en el voltaje $V_1$ como en $V_2$. Si por el contrario sumamos los elementos de cada fila, obtendremos la cantidad de pilas fabricadas por voltaje, de todos los tipos.

Nótese sin embargo que la suma de los elementos de la matriz $P$ no adquiere un significado relevante. Es decir, la información que proporcionan las matrices depende completamente del contexto al que se refieran.

**Definición 2.3**

Una matriz se dice *cuadrada* si $m = n$. El conjunto de matrices cuadradas de orden $n$ con coeficientes en $\mathbb{K}$ se notará por $\mathcal{M}_n(\mathbb{K})$.

Una matriz cuadrada se dice *simétrica* si $a_{ij} = a_{ji}$, $\forall i, j$.

Se denomina matriz *traspuesta* (o *transpuesta*) de $A = (a_{ij})$ (no necesariamente cuadrada) a la matriz $A^T = (a_{ji})$, es decir, la que resulta al intercambiar sus filas por sus columnas.

A partir de la definición se deduce que una matriz es simétrica si y solo si es igual a su traspuesta.

**Definición 2.4**

Los elementos $a_{ii}$, $1 \leq i \leq \text{mín}\{m, n\}$ son denominados *elementos diagonales*, y conforman la llamada *diagonal principal* de la matriz.

Una matriz con ceros por debajo de la diagonal principal se denomina

*triangular superior*. Si los ceros están por encima se dirá *triangular inferior*, y si están tanto encima como debajo se dirá matriz *diagonal*.

**Nota 2.1**

Obsérvese que la trasposición de una matriz corresponde a una "simetría" respecto de la diagonal principal. De este modo, una matriz simétrica es aquella que al trasponerla, no cambia.

**Ejemplo 2.2**

Las traspuestas de las matrices del ejemplo 2.1 son

$$F^T = \begin{pmatrix} 20 & 22 \\ 19 & 19 \\ 18 & 21 \end{pmatrix} \qquad P^T = \begin{pmatrix} 70 & 45 & 60 \\ 120 & 65 & 50 \end{pmatrix}$$

### Operaciones con matrices[2]

Las operaciones más simples que podemos realizar con matrices son la *suma* y el *producto por un escalar*, que se definen del siguiente modo:

**Definición 2.5**

Dadas $A, B \in \mathcal{M}_{m\times n}(\mathbb{K})$, con $A = (a_{ij})$ y $B = (b_{ij})$, y $\alpha \in \mathbb{K}$, se definen:

(I) la *suma* de $A$ y $B$ como la matriz $A + B \in \mathcal{M}_{m\times n}(\mathbb{K})$ donde $(A + B)_{ij} = (a_{ij} + b_{ij})$.

(II) el *producto por escalar* de $\alpha \in \mathbb{K}$ y $A$ como la matriz $\alpha A \in \mathcal{M}_{m\times n}(\mathbb{K})$ dada por $(\alpha A)_{ij} = \alpha a_{ij}$.

[2] Fueron introducidas por el matemático inglés Arthur Cayley en 1857.

Es decir, podemos sumar matrices siempre que ambas tengan el mismo orden, y la matriz suma se obtiene sumando los elementos que están en la misma posición. Por otro lado, el producto de una matriz por un escalar es otra matriz en la que cada elemento está multiplicado por ese número.

**Ejemplo 2.3**

Dadas las matrices

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathcal{M}_{3\times 2}(\mathbb{R})$$

entonces

$$A + B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 3 \end{pmatrix}, \qquad 2A = \begin{pmatrix} 2 & 0 \\ 2 & 2 \\ 0 & 4 \end{pmatrix}$$

Los siguientes dos resultados establecen propiedades elementales de la suma y del producto por un escalar. Las demostraciones de estos resultados son consecuencia inmediata de la definición y de las correspondientes propiedades del cuerpo $\mathbb{K}$.

**Proposición 2.1 (Propiedades de la suma de matrices)**

Si $A, B, C \in \mathcal{M}_{m\times n}(\mathbb{K})$, se verifican las siguientes propiedades:

(i) Conmutativa: $A + B = B + A$

(ii) Asociativa: $A + (B + C) = (A + B) + C$

(iii) Elemento neutro: existe una única matriz $\mathbf{0} \in \mathcal{M}_{m\times n}(\mathbb{K})$ tal que $A + \mathbf{0} = A$, $\forall A \in \mathcal{M}_{m\times n}(\mathbb{K})$, que se denomina *matriz nula* y cuyos elementos son todos cero.

(iv) Elemento opuesto o simétrico: para cada matriz $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ existe una única matriz $D$ tal que $A + D = \mathbf{0}$. Se notará $D = -A$.

En consecuencia, el par $(\mathcal{M}_{m\times n}(\mathbb{K}), +)$ es un grupo conmutativo (véase el Apéndice A).

**Proposición 2.2 (Propiedades del producto por un escalar)**

Si $\alpha, \beta \in \mathbb{K}$ y $A, B \in \mathcal{M}_{m\times n}(\mathbb{K})$ se verifican las siguientes propiedades:

(I) Pseudoasociativa: $\alpha(\beta A) = (\alpha\beta)A$

(II) Distributiva respecto de la suma de escalares: $(\alpha+\beta)A = \alpha A + \beta A$

(III) Distributiva respecto de la suma de matrices: $\alpha(A+B) = \alpha A + \alpha B$

Una vez vista la operación suma de matrices, parecería bastante natural definir el *producto de matrices* de forma análoga, es decir, multiplicando los elementos que se hayan en la misma posición. Lamentablemente, el producto de matrices no se define de ese modo, sino de una forma un poco más enrevesada:

**Definición 2.6**

Dadas dos matrices $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ y $B \in \mathcal{M}_{n\times p}(\mathbb{K})$, con $A = (a_{ik})$ y $B = (b_{kj})$ se define el *producto* de $A$ por $B$ como la matriz $AB \in \mathcal{M}_{m\times p}(\mathbb{K})$ cuyos elementos vienen dados por

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

La figura 2.1 ilustra el proceso de multiplicación. Para obtener el elemento $c_{ij}$ de la matriz producto $AB$ (el $c_{22}$ en la figura 2.1) debemos multiplicar cada uno de los elementos de la fila $i$ de la matriz $A$ (fila 2 en la figura 2.1) por el correspondiente elemento de la columna $j$ de la matriz $B$ (columna 2 en la figura 2.1), y calcular la suma de todos esos productos.

Por otra parte, hay que resaltar que para poder realizar el producto de dos matrices $A$ y $B$, los órdenes de las matrices deben guardar una relación: en concreto, el número de columnas de $A$ debe ser igual al número de filas de $B$ (para efectuar el producto $AB$).

**Ejemplo 2.4**

Multipliquemos las matrices $F$ y $P$ del ejemplo 2.1:

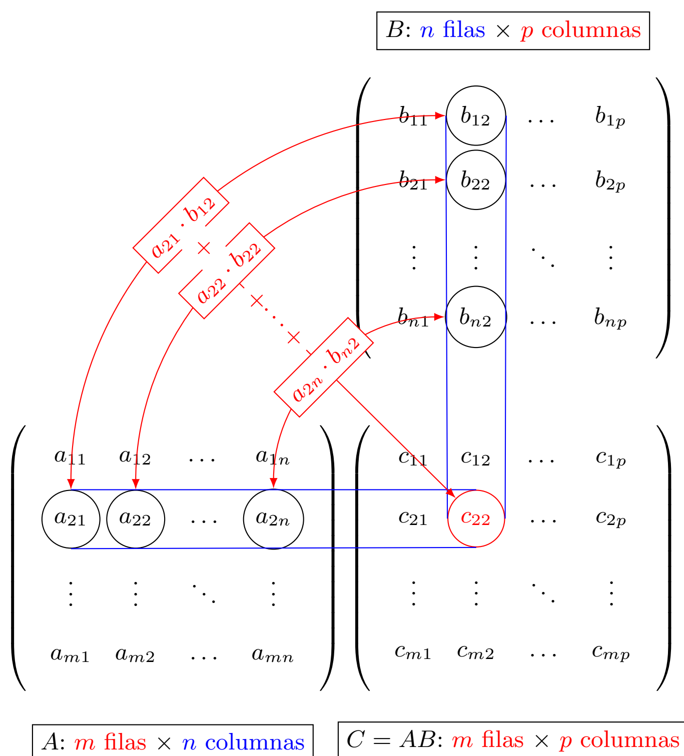



Figura 2.1: Ilustración del producto de matrices

$$FP = \begin{pmatrix} 20 \cdot 70 + 19 \cdot 45 + 18 \cdot 60 & 20 \cdot 120 + 19 \cdot 65 + 18 \cdot 50 \\ 22 \cdot 70 + 19 \cdot 45 + 21 \cdot 60 & 22 \cdot 120 + 19 \cdot 65 + 21 \cdot 50 \end{pmatrix}$$

$$= \begin{pmatrix} 3335 & 4535 \\ 3655 & 4925 \end{pmatrix}$$

¿Qué significado tienen las entradas de la diagonal de la matriz producto? Por ejemplo, el elemento $(FP)_{11}$ corresponde a la multiplicación del número de pilas fabricadas de los modelos de voltaje $V_1$ por su precio correspondiente, de modo que el valor 3335 nos da el beneficio por la venta de todas las pilas de voltaje $V_1$. Análogamente el elemento $(FP)_{22}$ nos proporciona el beneficio de la venta de todas las pilas de voltaje $V_2$.

¿Puede el lector averiguar el significado de los elementos diagonales de la matriz $PF$?

Es muy importante observar que **¡el producto de matrices no es conmutativo!** En algunos casos simplemente no es posible hacer el producto en ambos sentidos, pues los órdenes no coinciden. Cuando sí es posible hacer el producto en ambos sentidos, tampoco tiene por qué ser conmutativo, como se muestra en el ejemplo 2.5

**Ejemplo 2.5**

Dadas las matrices

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} \in \mathcal{M}_{3\times 2}(\mathbb{R}), \quad B = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \; C = \begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{R})$$

entonces

$$AB = \begin{pmatrix} 1 & 1 \\ 3 & 2 \\ 4 & 2 \end{pmatrix}, \quad BA \text{ no es posible}, \quad BC = \begin{pmatrix} 0 & 3 \\ 0 & 5 \end{pmatrix}, \quad CB = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

**Definición 2.7**

La matriz $I_n \in \mathcal{M}_n(\mathbb{K})$, dada por

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

o abreviadamente $(I_n)_{ij} = \delta_{ij}$, $1 \leq i,j \leq n$, donde $\delta_{ij}$ es el símbolo de Kronecker[3], se denomina *matriz identidad* de orden $n$.

[3] El símbolo o delta de Kronecker viene definido por

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

y le debe su nombre al matemático alemán Leopold Kronecker.

El producto de matrices verifica las siguientes propiedades:

**Proposición 2.3 (Propiedades del producto de matrices)**

Se verifican las siguientes propiedades (siempre que los productos sean posibles):

(I) Asociativa: $A(BC) = (AB)C$

(II) Distributiva respecto de la suma:

$$\begin{aligned} \text{(por la izquierda):} \quad & (B+C)A = BA + CA \\ \text{(por la derecha):} \quad & A(B+C) = AB + AC \end{aligned}$$

(III) Distributiva respecto del producto por escalar:

$$(\alpha A)B = A(\alpha B) = \alpha(AB)$$

(IV) Elemento neutro: se verifica $I_m A = A$, $AI_n = A$, $\forall A \in \mathcal{M}_{m\times n}(\mathbb{K})$, donde las matrices $I_n$, $I_m$ son las matrices identidad de órdenes $n$ y $m$, respectivamente.

La demostración de estas propiedades se basa exclusivamente en las definiciones de las operaciones correspondientes, aunque resulta un poco tediosa. A modo de ilustración, probaremos la asociatividad.

*Demostración:*

(I) Consideremos las matrices

$$A = (a_{ij})_{\substack{1\le i\le m,\\ 1\le j\le n}} \quad B = (b_{jk})_{\substack{1\le j\le n\\ 1\le k\le p}} \quad \text{y} \quad C = (c_{kl})_{\substack{1\le k\le p\\ 1\le l\le q}}$$

Nótese que los órdenes de las matrices deben ser los adecuados, pues en caso contrario no se podrían realizar los productos. Veamos que el elemento $il$ de la matriz $A(BC)$ es igual al elemento $il$ de la matriz $(AB)C$.

Observemos que el elemento $il$ de $A(BC)$ se obtiene al multiplicar la fila $i$ de $A$ por la columna $l$ de $BC$. Así pues, concentrémonos en $BC$. Sus elementos $(BC)_{jl}$ se obtienen (por definición) como:

$$(BC)_{jl} = \sum_{k=1}^{p} b_{ik} c_{kl}$$

Si ahora calculamos $A(BC)$,

$$((A(BC))_{il} = \sum_{j=1}^{n} a_{ij}(BC)_{jl} = \sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{p} b_{ik} c_{kl}\right) = \sum_{j=1}^{n}\sum_{k=1}^{p} a_{ij} b_{ik} c_{kl}$$

Ahora hagamos lo mismo con $(AB)C$. Los elementos de $AB$ son:

$$(AB)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

y por tanto:

$$((AB)C)_{il} = \sum_{k=1}^{p} (AB)_{ik} c_{kl} = \sum_{k=1}^{p} \left( \sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{kl} = \sum_{k=1}^{p} \sum_{j=1}^{n} a_{ij} b_{jk} c_{kl}$$

Puesto que los sumatorios son intercambiables, se obtiene el resultado.

■

Terminamos esta apartado sobre operaciones con matrices viendo cómo afecta la trasposición a la suma y el producto de matrices.

**Proposición 2.4 (Propiedades de la trasposición de matrices)**

Si $A, B$ son matrices y $\alpha$ un escalar, entonces

(I) $(A^T)^T = A$.

(II) $(\alpha A)^T = \alpha A^T$.

(III) $(A + B)^T = A^T + B^T$.

(IV) $(AB)^T = B^T A^T$.

*Demostración:*
La demostración de (I)–(III) es muy sencilla y se deja como ejercicio al lector. Para probar (IV) supongamos que $A$ es una matriz de orden $m \times n$ con elementos $(a_{ij})$ y $B$ una matriz de orden $n \times p$ con entradas $b_{jl}$. Denotando por $a'_{ji}$ y $b'_{lj}$ las entradas de $A^T$ y $B^T$, respectivamente, está claro que $a_{ij} = a'_{ji}$ y $b_{lj} = b'_{jl}$.

Si ahora denotamos por $c_{il}$ a las entradas de la matriz $AB$, entonces sabemos que

$$c_{il} = \sum_{j=1}^{n} a_{ij} b_{jl}$$

Denotando por $c'_{li}$ las entradas de $(AB)^T$, entonces

$$(AB)^T_{li} = c'_{li} = c_{il} = \sum_{j=1}^{n} a_{ij} b_{jl} = \sum_{j=1}^{n} b'_{lj} a'_{ji} = (B^T A^T)_{li}$$

■

2 2

## INVERSA DE UNA MATRIZ

Es sabido que el inverso de un escalar es otro escalar tal que, al multiplicar por éste, se obtiene el elemento neutro de la multiplicación, es decir, el uno. De forma similar podemos definir la inversa de una matriz, teniendo presente que el elemento neutro del producto es la matriz identidad (véase (IV) de la Proposición 2.3).

**Definición 2.8**

Dada una matriz cuadrada $A \in \mathcal{M}_n(\mathbb{K})$, se dice que la matriz $B \in \mathcal{M}_n(\mathbb{K})$ es su *inversa* si $AB = BA = I_n$, donde $I_n \in \mathcal{M}_n(\mathbb{K})$ es la matriz identidad. Se notará $B = A^{-1}$.

Si existe una tal matriz se dirá que $A$ es *regular* o *invertible*. En caso contrario se dirá que $A$ es *singular*.

Nótese que la inversa solo está definida para matrices cuadradas, y como muestra el siguiente ejemplo, no siempre existe.

**Ejemplo 2.6**

(I) Las matrices

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 2 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{pmatrix}$$

verifican que $AB = BA = I_2$ (el lector puede comprobarlo fácilmente), de manera que $B = A^{-1}$ (o también $A = B^{-1}$).

(II) Sin embargo, la matriz $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ no posee inversa, pues si existiera una tal matriz $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, se debería verificar

$$I_2 = AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$$

lo que es imposible.

**Teorema 2.1**

Si una matriz tiene inversa, entonces ésta es única.

*Demostración:*
En muchas ocasiones, para probar la unicidad de un objeto matemático, es habitual suponer que existen dos de tales objetos, y deducir, usando sus propiedades, que en realidad son iguales. De lo que se sigue que el objeto es único. Éste es el método que empleamos aquí: supongamos que $B$ y $C$ son dos inversas de $A$, es decir, $BA = AC = I_n$. Entonces,

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

Es decir, $B = C$, luego la inversa es única.

■

La siguiente proposición nos permite relajar ligeramente la definición de inversa, no siendo necesario comprobar la conmutatividad del producto.

**Proposición 2.5**

Si $A, B \in \mathcal{M}_n(\mathbb{K})$ y verifican que $AB = I_n$, entonces $BA = I_n$ y por tanto $B = A^{-1}$.

A pesar de que puede parecer evidente el resultado, la demostración requiere un cierto esfuerzo, por lo que la omitiremos.

El siguiente resultado nos muestra cómo funciona la inversión con el producto y la trasposición de matrices.

**Proposición 2.6**

Si $A, B \in \mathcal{M}_n(\mathbb{K})$, se tiene:

(I) Si $A$ y $B$ son invertibles, entonces $AB$ es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

(II) Si $A$ es invertible, entonces $A^T$ es invertible y $(A^T)^{-1} = (A^{-1})^T$.

*Demostración:*

(I) Bastará ver que $(AB)(B^{-1}A^{-1}) = I_n$, lo cual es evidente usando la propiedad asociativa del producto de matrices.

(II) Si $A$ es invertible, sabemos que $AA^{-1} = I_n$, de modo que calculando traspuestas

$$(AA^{-1})^T = I_n^T = I_n$$

Por otro lado, $(AA^{-1})^T = (A^{-1})^T A^T$ como vimos en (IV) de la Proposición 2.4. Es decir, el producto de $A^T$ con $(A^{-1})^T$ es la matriz identidad, por tanto una es la inversa de la otra.

■

El paso siguiente consiste en encontrar una forma de calcular inversas de matrices. El primer método que vamos a ver, el *método de Gauss*, probablemente es ya conocido por el lector. Para recordarlo, haremos una breve incursión en los sistemas de ecuaciones lineales, que retomaremos con mayor detenimiento en el tema siguiente.

## 2 3 SISTEMAS DE ECUACIONES LINEALES

Posiblemente, el ejemplo más habitual del uso de matrices está en la representación de *sistemas de ecuaciones lineales*, cuya definición damos a continuación.

**Definición 2.9**

Se denomina *sistema de ecuaciones lineal de $m$ ecuaciones con $n$ incógnitas* a una expresión del tipo siguiente:

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}\right\} \tag{2.1}$$

donde los $a_{ij}$ son denominados *coeficientes*, $x_i$ son las *incógnitas* y $b_i$ el *término independiente*.

Los sistemas de ecuaciones lineales aparecen frecuentemente en la resolución numérica de ecuaciones en derivadas parciales, en el planteamiento de problemas que provienen de la Física, la Economía, etc. En el tema 3 abordaremos cuestiones relacionadas con la resolución de este tipo de sistemas. De momento nos bastará con definir qué entendemos por solución de un sistema.

**Definición 2.10**

Se denomina *solución* del sistema (2.1) a toda $n$-upla $(\bar{x}_1, \dots, \bar{x}_n) \in \mathbb{K}^n$ que convierte la expresión (2.1) en una identidad.

Las matrices son muy adecuadas para simplificar la escritura de un sistema. Así, si consideramos las matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathcal{M}_{m\times n}(\mathbb{K}), \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{K}),$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathcal{M}_{m\times 1}(\mathbb{K})$$

el producto de matrices nos permite escribir el sistema de forma simple como $A\mathbf{x} = \mathbf{b}$.

El lector seguramente conocerá el *método de eliminación* habitualmente empleado para resolver sistemas lineales. El método está basado en las siguientes propiedades:

**Proposición 2.7**

Se verifican las siguientes propiedades:

(i) Si multiplicamos una ecuación por un escalar distinto de cero, las soluciones de (2.1) no varían.

(ii) Si intercambiamos dos ecuaciones en el sistema (2.1), las soluciones del mismo no cambian.

(iii) Si sustituimos una ecuación por el resultado de sumar o restar dicha ecuación con un múltiplo de otra, las soluciones del sistema (2.1) no varían.

La demostración es muy sencilla y se deja al lector.

Estas propiedades nos permiten poner en marcha un método simple y directo para resolver sistemas lineales conocido como *método de Gauss*.[4]

[4] Este método ya aparece en un libro de matemáticas chino titulado *Jiuzhang suanshu* o

### Método de Gauss

El objetivo de este método es la transformación del sistema original en uno más fácil de resolver mediante la aplicación de las propiedades descritas en la Proposición 2.7. Lo veremos directamente a través de un ejemplo.

**Ejemplo 2.7**

Resolver mediante el método de Gauss el siguiente sistema:

$$\left.\begin{array}{rcl} x+2y+3z &=& 7 \\ x-3y+2z &=& 5 \\ x+y+z &=& 3 \end{array}\right\}$$

Realizaremos las siguientes operaciones con las ecuaciones del sistema:

$$\left.\begin{array}{rcl} x+2y+3z &=& 7 \\ x-3y+2z &=& 5 \\ x+y+z &=& 3 \end{array}\right\} \xrightarrow[3^{\text{a}}-1^{\text{a}}]{2^{\text{a}}-1^{\text{a}}} \left.\begin{array}{rcr} x+2y+3z &=& 7 \\ -5y-z &=& -2 \\ -y-2z &=& -4 \end{array}\right\} \xrightarrow{1^{\text{a}}\leftrightarrow 2^{\text{a}}}$$

$$\left.\begin{array}{rcr} x+2y+3z &=& 7 \\ -y-2z &=& -4 \\ -5y-z &=& -2 \end{array}\right\} \xrightarrow{5\cdot 2^{\text{a}}-3^{\text{a}}} \left.\begin{array}{rcr} x+2y+3z &=& 7 \\ -y-2z &=& -4 \\ -9z &=& -18 \end{array}\right\}$$

En cada transformación del sistema hemos usado alguna de las propiedades de la Proposición 2.7. En concreto, en la primera transformación hemos sustituido la segunda ecuación por la diferencia entre la segunda y la primera ecuación, y luego hemos hecho lo propio entre la tercera y la primera. De esta forma hemos eliminado la variable $x$ en las dos últimas ecuaciones. En la segunda transformación simplemente hemos intercambiado las ecuaciones segunda y tercera con objeto de simplificar el siguiente paso. Finalmente hemos llegado a un sistema en el que han desaparecido las dos primeras variables de la última ecuación y la primera variable en la segunda ecuación. El sistema resultante (denominado *sistema triangular*) es muy fácil de resolver. Empezamos por resolver la última ecuación ($z=2$), y su valor lo sustituimos en la anterior, de manera que ésta es también fácil de resolver ($y=0$). Sucesivamente, sustituimos los valores obtenidos en la ecuación anterior para obtener la solución de la primera incógnita ($x=1$). A este método de resolver un sistema triangular se le llama una *subida*.

---

*Nueve capítulos del arte matemático* cuyo origen algunos historiadores marcan alrededor del año 200 a.C.

Como hemos visto, el método de Gauss consiste en "triangular" el sistema dado de manera que pueda resolverse fácilmente mediante una subida. Es importante observar que el papel de las incógnitas es superfluo. Únicamente son necesarios los coeficientes de las mismas en el sistema así como el término independiente, es decir, podemos trabajar fácilmente con matrices, como vemos en el siguiente ejemplo.

**Ejemplo 2.8**

Resolver el sistema mediante el método de Gauss:

$$\left.\begin{array}{rcr} x+3y+z & = & -3 \\ 3x+9y+4z & = & -7 \\ 2x-y+z & = & 6 \end{array}\right\}$$

En forma matricial,

$$\left(\begin{array}{rrr|r} 1 & 3 & 1 & -3 \\ 3 & 9 & 4 & -7 \\ 2 & -1 & 1 & 6 \end{array}\right) \xrightarrow[F_3-2F_1]{F_2-3F_1} \left(\begin{array}{rrr|r} 1 & 3 & 1 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & -7 & -1 & 12 \end{array}\right)$$

$$\xrightarrow{F_2 \leftrightarrow F_3} \left(\begin{array}{rrr|r} 1 & 3 & 1 & -3 \\ 0 & -7 & -1 & 12 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

Resolvemos mediante una subida y obtenemos $z = 2$, $y = -2$ y $x = 1$. La terna $(1, -2, 2)$ es solución de este sistema.

Es habitual marcar con una línea la separación entre la matriz de los coeficientes del sistema y los términos independientes. Por otra parte, nótese que hemos operado con las ecuaciones a través de las filas de la matriz, convenientemente notadas como $F_i$. Es muy aconsejable anotar en cada paso las operaciones realizadas.

Obsérvese además que la matriz resultante es una matriz triangular superior.

### Matrices elementales

Las operaciones llevadas a cabo en el método de Gauss se conocen como *operaciones elementales* y pueden ser interpretadas mediante multiplicación de *matrices elementales* que definimos a continuación:

**Definición 2.11**

(I) $E_{ij}$ es la matriz que se obtiene de la identidad cuando intercambiamos la fila $i$ con la fila $j$, es decir:

$$E_{ij} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \longleftarrow \text{Fila } i \\ \\ \longleftarrow \text{Fila } j \\ \\ \\ \end{matrix}$$

(II) $E_i(\lambda)$ es la matriz obtenida al multiplicar la fila $i$ por $\lambda$ en la matriz identidad, es decir:

$$E_i(\lambda) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \lambda & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ \longleftarrow \text{Fila } i \\ \\ \\ \\ \end{matrix}$$

(III) $E_{ij}(\lambda)$ es la matriz obtenida de la matriz identidad al sumar a la fila $j$, la fila $i$ previamente multiplicada por $\lambda$, esto es:

$$\begin{matrix} & & \text{Col. i} & & \text{Col. j} & & \\ & & \downarrow & & \downarrow & & \end{matrix}$$

$$E_{ij}(\lambda) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & \lambda & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \longleftarrow \text{Fila } i \\ \\ \longleftarrow \text{Fila } j \\ \\ \\ \end{matrix}$$

El siguiente resultado nos dice cómo funciona el método de Gauss a través de las operaciones con matrices elementales.

**Proposición 2.8**

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ y sean $E$ y $F$ matrices elementales de orden $m$ y $n$, respectivamente. Entonces el producto $EA$ es la matriz que se obtiene de $A$ cuando aplicamos a sus filas la misma transformación que aplicamos a la identidad para obtener $E$.

Análogamente, $AF$ es la matriz que se obtiene de $A$ cuando aplicamos a sus columnas la misma transformación que aplicamos a la identidad para obtener $F$.

*Demostración:*
Probemos solamente que el resultado es cierto para la matriz elemental $E = E_{ij}(\lambda)$. El resto de las pruebas es similar y se deja al lector.

En primer lugar observemos que si denotamos por $e_{kl}$ los coeficientes de la matriz $E$, entonces

$$e_{kl} = \begin{cases} 1 & \text{si } k = l, \\ \lambda & \text{si } k = j \text{ y } l = i, \\ 0 & \text{en el resto.} \end{cases} \qquad 1 \leq k, l \leq m$$

La matriz producto $EA$ tiene como elementos

$$(EA)_{kp} = \sum_{l=1}^{m} e_{kl} a_{lp}, \quad 1 \leq k \leq m, \ 1 \leq p \leq n$$

Si $k \neq j$, el único elemento $e_{kl}$ con $1 \leq l \leq m$ no nulo es $e_{kk}$ (es decir, el único elemento de la fila $k$ no nulo es el de la diagonal), luego $(EA)_{kp} = a_{kp}$. Esto es, todas las filas (salvo la j-ésima) de la matriz producto son iguales a $A$.

En la fila $j$, sólo hay dos elementos no nulos, $e_{ji} = \lambda$ y $e_{jj} = 1$, luego

$$(EA)_{jp} = \lambda a_{ip} + a_{jp}, \quad 1 \leq p \leq n$$

luego la fila $j$-ésima del producto $EA$ corresponde a la suma de los elementos de la fila $i$ de $A$ multiplicados por $\lambda$, más los elementos de la fila $j$. Esto prueba el resultado.

■

**Ejemplo 2.9**

Consideremos la matriz $A = \begin{pmatrix} -1 & 3 & 4 \\ 2 & 1 & 0 \\ 2 & -1 & -1 \end{pmatrix}$

(i) Multipliquemos por la izquierda por la matriz $E_3(2)$:

$$E_3(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \Rightarrow E_3(2)A = \begin{pmatrix} -1 & 3 & 4 \\ 2 & 1 & 0 \\ 4 & -2 & -2 \end{pmatrix}$$

La operación realizada equivale a multiplicar por 2 la tercera fila de $A$.

(ii) Multipliquemos a la izquierda por la matriz $E_{13}(-2)$:

$$E_{13}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \Rightarrow E_{13}(-2)A = \begin{pmatrix} -1 & 3 & 4 \\ 2 & 1 & 0 \\ 4 & -7 & -9 \end{pmatrix}$$

El producto es equivalente a multiplicar por $-2$ la primera fila de $A$ y sumarla a la tercera.

(iii) Multipliquemos a la izquierda por la matriz $E_{23}$:

$$E_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \Rightarrow E_{23}A = \begin{pmatrix} -1 & 3 & 4 \\ 2 & -1 & -1 \\ 2 & 1 & 0 \end{pmatrix}$$

El producto equivale a intercambiar entre sí las filas segunda y tercera.

El lector puede comprobar qué sucede cuando las multiplicaciones se realizan a la derecha.

Gracias a estas matrices elementales, es evidente que el método de Gauss no supone más que la multiplicación sucesiva de matrices elementales. Es por ello que ahora, más que nunca, se hace indispensable anotar las operaciones realizadas entre las filas de la matriz, pues cada una de ellas corresponderá a una multiplicación por una matriz elemental. En la sección 3.3.1 usaremos este hecho con más detalle para resolver sistemas de una forma particular.

2 4

## CÁLCULO DE LA INVERSA MEDIANTE OPERACIONES ELEMENTALES

En esta sección vamos a usar el método de Gauss para calcular la matriz inversa de una matriz cuadrada. Comencemos con un ejemplo sencillo.

**Ejemplo 2.10**

Consideremos la matriz

$$A = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$$

y calculemos su inversa.

Para ello debemos buscar una matriz $B \in \mathcal{M}_{2\times 2}$ tal que $AB = I_2$. Si ponemos,

$$B = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$$

entonces

$$AB = I_2 \Rightarrow \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

lo cual equivale a resolver el sistema

$$\left.\begin{array}{rcl} 2x_1 + x_3 & = & 1 \\ 2x_2 + x_3 & = & 0 \\ 4x_1 + 3x_3 & = & 0 \\ 4x_2 + 3x_4 & = & 1 \end{array}\right\}$$

Antes de resolver este sistema por el método de Gauss podemos observar que dicho sistema está *desacoplado*, es decir, podemos resolver independientemente las ecuaciones primera y tercera, por un lado, y la segunda y la cuarta por otro.

Resolviendo el primero de esto sistemas mediante Gauss se obtiene:

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 4 & 3 & 0 \end{array}\right) \xrightarrow{F_2 - 2F_1} \left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & 1 & -2 \end{array}\right)$$

de donde $x_3 = -2$ y $x_1 = \frac{3}{2}$, después de resolver mediante una subida. No obstante, existe una alternativa a la resolución mediante una subida para resolver el sistema: podemos *diagonalizar* la matriz de los coeficientes siguiendo el esquema descrito por el método de Gauss, es decir, haciendo ceros ahora por

encima de la diagonal principal:

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & 1 & -2 \end{array}\right) \xrightarrow{F_1-F_2} \left(\begin{array}{cc|c} 2 & 0 & 3 \\ 0 & 1 & -2 \end{array}\right) \xrightarrow{\frac{1}{2}F_1} \left(\begin{array}{cc|c} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -2 \end{array}\right)$$

De hecho, hemos dado un pequeño paso más; una vez diagonalizada la matriz hemos multiplicamos adecuadamente cada fila para que en la diagonal de la matriz de los coeficientes obtengamos un 1. De este modo se observa que el término independiente muestra exactamente la solución de sistema, mientras que la matriz de los coeficientes se ha transformado en la matriz identidad. Esta "extensión" del método de Gauss es conocida como *método de Gauss-Jacobi*.

Procedemos del mismo modo con el otro par de ecuaciones para encontrar $x_2$ y $x_4$:

$$\left(\begin{array}{cc|c} 2 & 1 & 0 \\ 4 & 3 & 1 \end{array}\right) \xrightarrow{F_2-2F_1} \left(\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right) \xrightarrow{F_1-F_2} \left(\begin{array}{cc|c} 2 & 0 & -1 \\ 0 & 1 & 1 \end{array}\right)$$
$$\xrightarrow{\frac{1}{2}F_1} \left(\begin{array}{cc|c} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 1 \end{array}\right)$$

de donde deducimos que $x_2 = -\frac{1}{2}$ y $x_4 = 1$. Así pues, la inversa de $A$ viene dada por

$$A^{-1} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -2 & 1 \end{pmatrix}$$

No obstante, es interesante observar que los dos sistemas desacoplados, cuya solución nos ha permitido obtener la matriz inversa, tienen la misma matriz de coeficientes (igual a la matriz $A$), por lo que la aplicación del método de Gauss-Jacobi realiza exactamente las mismas operaciones en ambos. Por tanto, es posible resolver ambos sistemas simultáneamente si escribimos juntos los dos términos independientes:

$$\left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 4 & 3 & 0 & 1 \end{array}\right) \xrightarrow[\frac{1}{2}F_1]{\substack{F_2-2F_2 \\ F_1-F_2}} \left(\begin{array}{cc|cc} 1 & 0 & \frac{3}{2} & -\frac{1}{2} \\ 0 & 1 & -2 & 1 \end{array}\right)$$

Vemos que la matriz de inicio esta formada por la "unión" de la matriz $A$ e $I_2$, mientras que la matriz final está formada por $I_2$ y $A^{-1}$. Dicho de otro modo, las operaciones elementales que nos conducen desde la matriz $A$ a la matriz identidad, son las que nos llevan desde la matriz identidad a $A^{-1}$.

Como hemos podido ver en este ejemplo, el cálculo de la inversa de una matriz $A \in \mathcal{M}_n(\mathbb{K})$ por el método de Gauss consiste en diagonalizar la matriz $(A \mid I_n)$ hasta conseguir la identidad mediante transformaciones elementales. La matriz resultante en la derecha es la inversa de la dada.

**Ejemplo 2.11**

Encontrar la inversa de la matriz

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 3 \\ 2 & -1 & -8 \end{pmatrix}$$

Escribimos la matriz junto con la identidad de orden 3:

$$A = \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 2 & -1 & -8 & 0 & 0 & 1 \end{array}\right) \xrightarrow{F_3-2F_1} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & -5 & -8 & -2 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{F_3+5F_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 7 & -2 & 5 & 1 \end{array}\right) \xrightarrow{\frac{1}{7}F_3} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \end{array}\right)$$

$$\xrightarrow{F_2-3F_3} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{6}{7} & -\frac{8}{7} & -\frac{3}{7} \\ 0 & 0 & 1 & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \end{array}\right)$$

$$\xrightarrow{F_1-2F_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{5}{7} & \frac{16}{7} & \frac{6}{7} \\ 0 & 1 & 0 & \frac{6}{7} & -\frac{8}{7} & -\frac{3}{7} \\ 0 & 0 & 1 & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \end{array}\right)$$

Luego,

$$A^{-1} = \begin{pmatrix} -\frac{5}{7} & \frac{16}{7} & \frac{6}{7} \\ \frac{6}{7} & -\frac{8}{7} & -\frac{3}{7} \\ -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \end{pmatrix}$$

El lector puede comprobar que en efecto $AA^{-1} = I_3$.

**Ejemplo 2.12**

Calcular la inversa de

$$A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}$$

Procediendo como en el ejemplo anterior:

$$\left(\begin{array}{cc|cc} 2 & 4 & 1 & 0 \\ 3 & 6 & 0 & 1 \end{array}\right) \xrightarrow{\frac{1}{2}F_1} \left(\begin{array}{cc|cc} 1 & 2 & \frac{1}{2} & 0 \\ 3 & 6 & 0 & 1 \end{array}\right) \xrightarrow{F_2-3F_1} \left(\begin{array}{cc|cc} 1 & 2 & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{3}{2} & 1 \end{array}\right)$$

Podemos observar en este caso que no hay forma alguna de obtener la matriz identidad mediante transformaciones elementales de $A$. Esto significa que no existe $A^{-1}$.

En definitiva, mediante el método de Gauss tenemos un método para, mediante operaciones elementales, obtener una matriz triangular superior. Si los elementos diagonales de esta matriz son todos distintos de cero, es posible realizar nuevas operaciones que nos conduzcan a una matriz diagonal, y por tanto, aplicando (i) de la Proposición 2.7 llegar a la matriz identidad mediante operaciones elementales.

Puesto que la aplicación de operaciones elementales equivalen a multiplicar por matrices elementales, podemos escribir:

$$E_k \cdots E_2 E_1 A = I \Rightarrow A = (E_k \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$

donde $E_1, \ldots, E_k$ son las matrices elementales correspondientes a las operaciones realizadas. Por ejemplo, en el ejemplo 2.11, las matrices elementales usadas son:

$$E_1 = E_{13}(-2),\ E_2 = E_{23}(5),\ E_3 = E_3(\tfrac{1}{7}),\ E_4 = E_{32}(-3),\ E_5 = E_{21}(-2)$$

Es fácil demostrar que las matrices elementales son invertibles, y su inversa es un matriz elemental.[5] De este modo se tiene el siguiente resultado.

**Teorema 2.2**

Una matriz es invertible si y sólo si es producto de matrices elementales.

[5] Basta observar que para llegar a la identidad desde una matriz elemental debemos realizar la operación elemental “inversa”.

**Nota 2.2**

Como ya hemos visto, la definición del producto de matrices nos permite ver el sistema (2.1) en forma matricial como $A\mathbf{x} = \mathbf{b}$. Si resulta que este sistema es *cuadrado*, esto es, tiene el mismo número de ecuaciones que de incógnitas ($m = n$), y además la matriz de los coeficientes es invertible, multiplicando por $A^{-1}$ a la izquierda obtenemos $A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$. Puesto que $A^{-1}A = I_n$ e $I_n\mathbf{x} = \mathbf{x}$, se sigue que $\mathbf{x} = A^{-1}\mathbf{b}$.

## 2 5 DETERMINANTE DE UNA MATRIZ

El concepto de determinante[6] se va a definir de manera inductiva. Para una visión general del método de inducción matemática, que será usado con frecuencia en las demostraciones de esta sección, nos remitimos al apéndice A.

Comenzaremos con un par de definiciones previas.

**Definición 2.12**

Se denomina *submatriz* de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ dada, a cualquier matriz obtenida de $A$ eliminando un conjunto determinado de filas y/o columnas.

**Definición 2.13**

Se denomina *matriz adjunta del elemento ij* de una matriz $A$, y se denotará por $A_{ij}$, a la submatriz obtenida al eliminar la fila $i$ y la columna $j$ de dicha matriz.

Nótese que la matriz adjunta de un elemento cualquiera tiene siempre un orden menos que la matriz inicial (véase la figura 2.2).

[6] Los determinantes aparecen con anterioridad a las matrices. El de orden dos ya aparece en la obra de Cardano en 1545, en relación con la solución de sistemas lineales de dos ecuaciones con dos incógnitas. Los de orden superior aparecen casi simultáneamente en las obras del japonés Kowa Seki y el alemán Gottfried Wilhelm von Leibniz alrededor del año 1680. Pero es el francés Augustin Louis Cauchy el que en 1812 comienza a usar los determinantes en el sentido actual, probando varias de sus propiedades.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \longrightarrow A_{22} = \begin{pmatrix} a_{11} & a_{13} & \cdots & a_{1n} \\ a_{31} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

Figura 2.2: Matriz adjunta del elemento 2 2

**Definición 2.14**

Sea $A \in \mathcal{M}_n(\mathbb{K})$. Se define el *determinante* de $A$, que denotaremos por $\det A$ ó $|A|$, del siguiente modo:

- Si $n = 1$, $|A| = a_{11}$.
- Si $n > 1$,

$$|A| = a_{11}|A_{11}| - a_{21}|A_{21}| + \cdots + (-1)^{n+1}a_{n1}|A_{n1}| = \sum_{i=1}^{n}(-1)^{i+1}a_{i1}|A_{i1}|$$

  donde $|A_{i1}|$ es el determinante de la matriz adjunta del elemento $i1$, que corresponde a una matriz cuadrada de orden $n-1$.

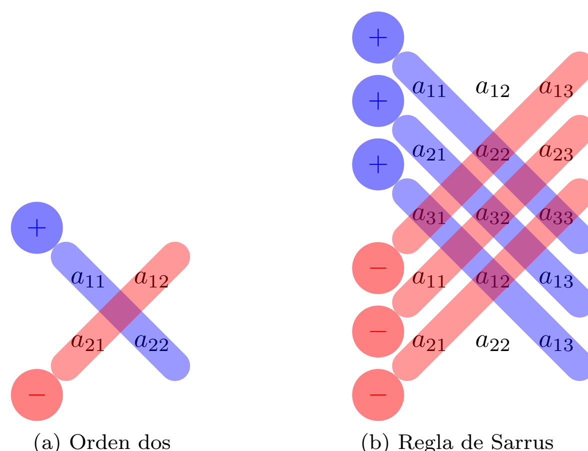
Veamos cómo desarrollar la definición de determinante en los casos $n = 2$ y $n = 3$:

(I) Si $n = 2$, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, entonces

$$|A| = a_{11}|A_{11}| - a_{21}|A_{21}| = a_{11}a_{22} - a_{21}a_{12}$$

(II) Si $n = 3$, $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, entonces

$$\begin{aligned} |A| &= a_{11}|A_{11}| - a_{21}|A_{21}| + a_{31}|A_{31}| = \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} \end{aligned}$$




(a) Orden dos (b) Regla de Sarrus

Figura 2.3: Fórmulas para los determinantes de orden dos y tres

$$+a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}$$

Como se puede observar, el determinante de una matriz es un número que se calcula mediante sumas y productos de sus coeficientes. El determinante de una matriz de orden dos tiene una expresión fácil de recordar, pues se trata de calcular el producto de los elementos de la diagonal principal y restarle el producto de los elementos de la diagonal opuesta (véase la figura 2.3a)

Sin embargo, para el determinante de orden tres aparecen seis productos, que dificultan la memorización de la fórmula que proporciona su valor. No obstante, existe una sencilla regla mnemotécnica para recordar la expresión del determinante de una matriz de orden tres, la conocida como *regla de Sarrus*,[7] que nos indica cuáles son los productos con signo positivo y cuáles los de signo negativo (véase la figura 2.3b).

Es importante resaltar que estas reglas (que parecen tener un cierto parecido) no son extensibles a determinantes de orden superior. En particular, ¡la regla de Sarrus sólo es válida para determinantes de orden tres! Por otra parte, aunque la regla tiene un cierto "encanto" que la hace muy popular entre los estudiantes, veremos más adelante que es preferible calcular determinantes de orden tres empleando alguna de las propiedades que siguen. Al fin y al cabo, la regla de Sarrus obliga al cálculo de seis productos, que pueden hacerse un tanto engorrosos.

El resto de la sección está dedicada a presentar una serie de propiedades que nos permitirán simplificar el cálculo de determinantes.

[7]Introducida por el matemático francés Pierre Frédéric Sarrus en 1833.

### 2.5.1 Propiedades de los determinantes

**Proposición 2.9**

Si una matriz cuadrada $A$ tiene una fila completa de ceros, entonces su determinante es nulo.

*Demostración:*
Procederemos por inducción en el orden de la matriz. Si $n = 1$ el resultado es evidente. Supongamos ahora que el resultado es cierto para las matrices de orden $n-1$, y probémoslo para orden $n$. Supongamos además que la fila $j$-ésima es nula, es decir $a_{j1} = \cdots = a_{jn} = 0$. Entonces, por definición,

$$|A| = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} |A_{i1}|$$

Ahora bien, cada matriz $A_{i1}$ tiene orden $n-1$ (pues es la adjunta de una matriz de orden $n$) y todas, salvo $A_{j1}$, tienen una fila completa de ceros. Por hipótesis de inducción, $|A_{i1}| = 0\ \forall i \neq j$. Luego

$$|A| = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} |A_{i1}| = (-1)^{j+1} a_{j1} |A_{j1}|$$

pero este último sumando también es nulo, pues $a_{j1} = 0$ (ya que está en la fila $j$); de aquí se tiene el resultado.

■

**Proposición 2.10**

Se verifica:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1}+b_{i1} & \cdots & a_{in}+b_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ b_{i1} & \cdots & b_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \tag{2.2}$$

*Demostración:*
Nuevamente procedemos por inducción en $n$. Al igual que antes, para $n = 1$ el

resultado es trivial. Supongamos ahora el resultado cierto para el orden $n-1$ y probémoslo para $n$. Denotaremos por $C$ a la matriz del primer miembro de la igualdad en (2.2) y por $A$ y $B$ a las de la derecha, respectivamente. Queremos probar que $|C| = |A| + |B|$.

Por definición,

$$|C| = a_{11}|C_{11}| + \cdots + (-1)^{i+1}(a_{i1} + b_{i1})|C_{i1}| + \cdots + (-1)^{n+1}a_{n1}|C_{n1}|$$

observemos ahora que cada matriz $C_{j1}$ con $j \neq i$ es de orden $n-1$ y mantiene la misma estructura que $C$, por tanto podemos usar la hipótesis de inducción y asegurar que $|C_{j1}| = |A_{j1}| + |B_{j1}|$, si $j \neq i$. Así pues,

$$|C| = a_{11}(|A_{11}| + |B_{11}|) + \cdots + (-1)^{i+1}(a_{i1} + b_{i1})|C_{i1}| \\ + \cdots + (-1)^{n+1}a_{n1}(|A_{n1}| + |B_{n1}|)$$

Finalmente, basta observar que $C_{i1} = A_{i1} = B_{i1}$ (pues todas las matrices tienen las mismas filas excepto la $i$-ésima) y aplicar la definición de determinante para obtener lo deseado.

■

**Proposición 2.11**

Si $B$ es la matriz obtenida al multiplicar los elementos de una fila de la matriz $A$ por un escalar $\lambda \in \mathbb{K}$, entonces $|B| = \lambda|A|$.

*Demostración:*

Nuevamente el resultado es trivial para $n = 1$. Para probar el caso $n$, supuesto que se satisface el caso $n-1$, vemos que si

$$A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \quad \text{entonces} \quad B = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{i1} & \cdots & \lambda a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

Por tanto,

$$|B| = a_{11}|B_{11}| + \cdots + (-1)^{i+1}\lambda a_{i1}|B_{i1}| + \cdots + (-1)^{n+1}a_{n1}|B_{n1}|$$

Como en la demostración anterior, cada matriz $B_{j1}$, con $j \neq i$, tiene orden $n-1$ y la misma estructura que $B$. Por hipótesis de inducción, $|B_{j1}| = \lambda|A_{j1}|$.

Por otra parte, está claro que $B_{i1} = A_{i1}$, así pues, sustituyendo en la ecuación anterior

$$|B| = a_{11}\lambda|A_{11}| + \cdots + (-1)^{i+1}\lambda a_{i1}|A_{i1}| + \cdots + (-1)^{n+1}a_{n1}\lambda|A_{n1}| = \lambda|A|$$

■

**Proposición 2.12**

Si $B$ es la matriz que se obtiene de $A$ intercambiando dos de sus filas, entonces $|B| = -|A|$.

*Demostración:*
Como antes, procederemos por inducción en el orden de la matriz. Comenzamos probando el caso $n = 2$ (pues el caso $n = 1$ no adquiere significado en esta situación). Es claro que

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} = -(a_{21}a_{12} - a_{11}a_{22}) = -\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix}$$

Para probar el caso $n$, supuesto que se verifica el caso $n-1$, empezaremos probando que el resultado es cierto si intercambiamos dos filas consecutivas.

Sea $A = (a_{ij})$ y sea $B$ la matriz obtenida al intercambiar dos filas consecutivas, $i-1$ e $i$. Entonces:

$$B = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ a_{i-1,1} & \cdots & a_{i-1,n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \begin{matrix} \\ \\ \longleftarrow \text{Fila } i-1 \\ \longleftarrow \text{Fila } i \\ \\ \\ \end{matrix}$$

Usando la definición de determinante,

$$|B| = a_{11}|B_{11}| + \cdots + (-1)^{i}a_{i1}|B_{i-1,1}| + (-1)^{i+1}a_{i-1,1}|B_{i1}| + \cdots + (-1)^{n+1}a_{n1}|B_{n1}| \tag{2.3}$$

Ahora usamos la hipótesis de inducción en cada matriz $B_{j1}$, $j \neq i-1, i$. Puesto que son matrices de orden $n-1$ obtenidas al intercambiar dos filas consecutivas de las correspondientes matrices $A_{j1}$, entonces $|B_{j1}| = -|A_{j1}|$, si $j \neq i-1, i$.

Por otro lado, es fácil darse cuenta que $B_{i-1,1} = A_{i1}$ y que $B_{i1} = A_{i-1,1}$. Finalmente, sustituyendo en (2.3)

$$|B| = -a_{11}|A_{11}| + \cdots + (-1)^i a_{i1}|A_{i1}| + (-1)^{i+1} a_{i-1,1}|A_{i-1,1}| + \cdots + (-1)^{n+1} a_{n1}|A_{n1}| = -|A|$$

Para probar el caso general, es decir, cuando intercambiamos dos filas no necesariamente consecutivas, es suficiente observar que el intercambio de la fila $i$ con la fila $j$ equivale a realizar $2k-1$ cambios consecutivos, donde $k = |i-j|$. Por tanto, el signo se altera $2k-1$ veces, que al ser un número impar, supone un cambio de signo en el valor del determinante.

■

**Proposición 2.13**

Si una matriz tiene dos filas iguales su determinante es nulo.

*Demostración:*
La demostración es consecuencia inmediata de la proposición anterior, puesto que si intercambiamos dos filas iguales, el determinante de la matriz debe cambiar de signo, aunque la matriz es la misma, por lo que $|A| = -|A| \Rightarrow |A| = 0$.

■

El siguiente resultado es consecuencia de las proposiones anteriores y jugará un papel fundamental en el cálculo de determinantes.

**Proposición 2.14**

Si $B$ es la matriz obtenida de $A$ al sumar a una fila de $A$, un múltiplo de otra fila de la matriz, entonces $|B| = |A|$.

*Demostración:*
En efecto, si $B$ es de la forma especificada en el enunciado:

$$|B| = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{i1} + \lambda a_{j1} & \cdots & a_{in} + \lambda a_{jn} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

podemos hacer uso de las Proposiciones 2.10 y 2.11 para descomponer el determinante del siguiente modo:

$$|B| = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} + \lambda \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

y este último determinante es nulo puesto que tiene dos filas iguales.

■

Nótese que la transformación efectuada en esta proposición es una de las transformaciones típicas que se llevan a cabo en el método de Gauss, y que no altera el valor del determinante.

El resultado que sigue a continuación nos da un caracterización inmediata de las matrices regulares, esto es, las que poseen inversa.

**Proposición 2.15**

$A$ es una matriz regular si y sólo si $|A| \neq 0$.

*Demostración:*

Observemos en primer lugar que las Proposiciones 2.11, 2.12 y 2.14 afirman que los determinantes de las matrices elementales son $|E_{ij}| = -1$, $|E_i(\lambda)| = \lambda$ y $|E_{ij}(\lambda)| = 1$, pues tales matrices han sido obtenidas de la identidad mediante las operaciones señaladas en los resultados mencionados, y es fácil ver que el determinante de la matriz identidad (de cualquier orden) es 1. Por tanto, se deduce que

$$\det(EA) = \det(E)\det(A)$$

para cualquier matriz elemental $E$ y cualquier matriz $A$. Luego

$$\det(E_1 \cdots E_k A) = \det(E_1) \cdots \det(E_k)\det(A)$$

para cualesquiera matrices elementales $E_1, \ldots, E_k$. Como por el Teorema 2.2, $A$ es regular si y sólo si es producto de matrices elementales, es decir, $A = E_1' \cdots E_k'$, se tendrá que

$$\det(A) = \det(E_1' \cdots E_k') = \det(E_1') \cdots \det(E_k') \neq 0$$

■

**Proposición 2.16**

Si $A$, $B \in \mathcal{M}_n(\mathbb{K})$ entonces $\det(AB) = \det(A)\det(B)$.

*Demostración:*
Si $\det(A) = 0$ ó $\det(B) = 0$ entonces $A$ ó $B$ es singular, y por tanto también lo es $AB$ (véase Proposición 2.6), luego $\det(AB) = 0$ (por la Proposición 2.15).

Si $\det(A)$ y $\det(B)$ no son nulos, ambas son matrices invertibles y por tanto serán producto de matrices elementales. La demostración de la Proposición 2.15 nos da el resultado.

■

**Proposición 2.17**

$\det(A) = \det(A^T)$

El resultado se puede demostrar por inducción y se deja al lector (ejercicio 27).

**Nota 2.3**

Como consecuencia de este último resultado, las proposiciones 2.9–2.14, que enuncian propiedades por filas, pueden ser también enunciadas por columnas.

Finalmente, usando la nota 2.3 y la proposición 2.12 se tiene el siguiente resultado:

**Teorema 2.3 (Desarrollo por filas o columnas)**

Si $A = (a_{ij})$ entonces

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

y también

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

En definitiva, podemos desarrollar el determinante mediante los adjuntos de cualquier fila o columna.

En el siguiente ejemplo veremos como aplicar algunas de las propiedades anteriores para obtener el valor de los determinantes.

**Ejemplo 2.13**

Calcular el valor de los siguientes determinantes:

$$\text{(i)} \begin{vmatrix} 3 & 5 & 1 \\ 15 & 25 & 5 \\ 1 & 0 & 7 \end{vmatrix} \qquad \text{(ii)} \begin{vmatrix} 2 & 4 & 1 \\ 7 & 5 & 2 \\ 11 & 13 & 4 \end{vmatrix} \qquad \text{(iii)} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{vmatrix}$$

(i) Observemos que

$$\begin{vmatrix} 3 & 5 & 1 \\ 15 & 25 & 5 \\ 1 & 0 & 7 \end{vmatrix} = 5 \cdot \begin{vmatrix} 3 & 5 & 1 \\ 3 & 5 & 1 \\ 1 & 0 & 7 \end{vmatrix} = 5 \cdot 0 = 0$$

donde hemos aplicado la Proposición 2.11 a la segunda fila y luego la Proposición 2.13 a las filas uno y dos. En consecuencia, es inmediato observar que si una matriz tiene una fila (o columna) múltiplo de otra, entonces su determinante es cero.

(ii) Si aplicamos la Proposición 2.14 a las filas dos y tres,

$$\begin{vmatrix} 2 & 4 & 1 \\ 7 & 5 & 2 \\ 11 & 13 & 4 \end{vmatrix} \overset{F_3 - F_2}{=} \begin{vmatrix} 2 & 4 & 1 \\ 7 & 5 & 2 \\ 4 & 8 & 2 \end{vmatrix}$$

vemos que el determinante resultante tiene las filas una y dos proporcionales, luego por el comentario anterior el determinante es nulo.

(iii) Aplicando la Proposición 2.14 obtenemos un determinante con dos columnas iguales,

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{vmatrix} \overset{\substack{C_2 - C_1 \\ C_4 - C_3}}{=} \begin{vmatrix} 1 & 1 & 3 & 1 \\ 5 & 1 & 7 & 1 \\ 9 & 1 & 11 & 1 \\ 13 & 1 & 15 & 1 \end{vmatrix} = 0$$

En todos los casos anteriores hemos llegado a determinantes que tienen dos filas o columnas idénticas, lo que hace que el valor del determinante sea nulo. Obviamente, eso no va a ocurrir en general, pero en ocasiones merece la pena perder un poco de tiempo realizando operaciones sencillas en el determinante para ver si es posible obtener este hecho. Cuando esto no ocurre, la mejor forma de calcular un determinante es seguir los pasos que comentamos en el siguiente ejemplo.

**Ejemplo 2.14**

Calcular el valor del determinante:

$$\begin{vmatrix} -1 & 4 & 3 & -2 \\ -3 & 1 & 0 & 5 \\ 2 & 1 & -1 & 3 \\ 2 & 1 & 1 & 1 \end{vmatrix}$$

En lugar de proceder directamente desarrollando por una fila o columna haremos uso de los resultados anteriores (esencialmente la Proposición 2.14) a fin de lograr el mayor número de ceros en una misma fila o columna, de manera que al usar el Teorema 2.3, tengamos un número elevado de sumandos que quedarán multiplicados por cero. Así pues, trataremos de hacer ceros en la tercera columna (pues ya hay uno y los cálculos son más sencillos):

$$\begin{vmatrix} -1 & 4 & 3 & -2 \\ -3 & 1 & 0 & 5 \\ 2 & 1 & -1 & 3 \\ 2 & 1 & 1 & 1 \end{vmatrix} \overset{\substack{F_3+F_4 \\ F_1-3F_4}}{=} \begin{vmatrix} -7 & 1 & 0 & -5 \\ -3 & 1 & 0 & 5 \\ 4 & 2 & 0 & 4 \\ 2 & 1 & 1 & 1 \end{vmatrix}$$

desarrollando por la tercera columna

$$= -\begin{vmatrix} -7 & 1 & 5 \\ -3 & 1 & 5 \\ 4 & 2 & 4 \end{vmatrix} \overset{\substack{F_2-F_1 \\ F_3-2F_1}}{=} \begin{vmatrix} -7 & 1 & 5 \\ 4 & 0 & 10 \\ 18 & 0 & 14 \end{vmatrix} = \begin{vmatrix} 4 & 10 \\ 18 & 14 \end{vmatrix} = 2\cdot 2\begin{vmatrix} 2 & 5 \\ 9 & 7 \end{vmatrix} = -124.$$

**Nota 2.4**

Es muy importante observar que en todas las transformaciones que hemos realizado en el ejemplo anterior, la fila o columna que transformamos no se multiplica por ningún número, pues de hacerlo, el determinante quedaría multiplicado por dicho número.

En las dos siguientes secciones usaremos el cálculo de determinantes para obtener el rango de una matriz y su matriz inversa.

## 2 6 RANGO DE UNA MATRIZ

Una de las principales aplicaciones de los determinantes reside en el cálculo de rangos,[8] que será de mucha utilidad a lo largo de todo este texto. Es por ello que el método que vamos a ver en esta sección adquiere especial importancia.

**Definición 2.15**

Se denomina *menor de orden* $k$ de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ al determinante de una submatriz cuadrada de $A$ de orden $k$.

**Definición 2.16**

Se llama *rango* de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ al mayor orden de entre todos los menores de $A$ no nulos.

Es decir, si $\operatorname{rango}(A) = r$, toda submatriz cuadrada de $A$ de orden mayor que $r$ tiene determinante nulo, y al menos existe una submatriz cuadrada de orden $r$ con determinante no nulo. Nótese que es inmediato que si $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ entonces $\operatorname{rango}(A) \leq \min\{m, n\}$

**Teorema 2.4**

El rango de una matriz no se altera si:

(i) Intercambiamos entre sí dos filas (o dos columnas).

[8] Definido por primera vez por el matemático alemán Ferdinand Georg Frobenius en 1878.

(II) Trasponemos una fila con una columna (esto es, intercambiar la fila $i$ con la columna $i$).

(III) Multiplicamos los elementos de una fila (o columna) por un escalar distinto de cero.

(IV) Sumamos a una fila (o columna) los elementos de otra fila (o columna) multiplicados por un escalar.

La demostración de este resultado se deduce directamente de las propiedades de los determinantes.

**Cálculo del rango de una matriz: método del orlado.**

La definición de rango nos indica una forma de proceder con su cálculo: encontrar el mayor orden de entre todos los menores no nulos. Para ello, podríamos considerar todos los menores de una matriz y calcular sus determinantes, sin embargo, el siguiente método, conocido como *método del orlado*, nos proporciona un procedimiento en el que no es necesario obtener todos los menores de la matriz.

En primer lugar denominaremos menor *básico* a cualquier menor distinto de cero. A las filas y columnas de $A$ que conforman ese menor se les denomina filas y columnas básicas.

Para calcular el rango de una matriz se procede como sigue:

(I) Se suprimen todas las filas o columnas completamente nulas. Si todas las filas de una matriz son nulas entonces $\text{rango}(A) = 0$.

(II) Tomamos una fila donde exista un menor de orden uno no nulo. La fila y columna correspondientes a ese menor se consideran como básicas.[9]

(III) *Orlamos*, esto es, añadimos a la fila y columna básicas una nueva fila y una nueva columna, para formar un menor de orden dos. Si este menor es nulo, orlamos con la misma fila y una nueva columna, hasta encontrar un menor de orden dos distinto de cero.

Si todos los menores de orden dos obtenidos al orlar con la misma fila y todas las posibles columnas son nulos, podemos eliminar la fila en cuestión (puesto que no aporta rango a la matriz), y proceder con la siguiente, repitiendo el mismo procedimiento.

[9] Obviamente, si se considera cualquier otro menor de orden uno distinto de cero, las filas y columnas básicas serían otras. El procedimiento que lleva a cabo el método del orlado no es único.

(IV) Una vez encontrado un menor de orden dos distinto de cero, consideramos como básicas las filas y columnas que la componen y orlamos para encontrar un menor de orden tres no nulo, repitiendo el procedimiento anterior.

(V) Si al orlar un menor de orden $k$ con todas las filas y columnas disponibles no encontramos un menor de orden $k+1$ distinto de cero, entonces $\text{rango}(A) = k$.

**Ejemplo 2.15**

Calcular el rango de la matriz

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 3 \\ 0 & -3 & 0 & 4 & 1 \\ 0 & 2 & 0 & -5 & -4 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 \end{pmatrix} \in \mathcal{M}_{6\times 5}(\mathbb{R})$$

Está claro que el rango máximo de esta matriz es 5, puesto que no es posible construir un menor de orden 6. Puesto que la cuarta fila es completa de ceros, podemos eliminarla y proseguir sin ella. En la matriz resultante buscamos un menor de orden uno distinto de cero; por ejemplo, el formado por la fila 1, columna 2, que a partir de ahora, son consideradas básicas.

$$\begin{pmatrix} 0 & \boxed{1} & 0 & 1 & 3 \\ 0 & -3 & 0 & 4 & 1 \\ 0 & 2 & 0 & -5 & -4 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 \end{pmatrix} \in \mathcal{M}_{5\times 5}(\mathbb{R})$$

Orlamos este menor hasta encontrar uno distinto de cero. Es fácil ver que orlando con la segunda fila y la cuarta columna, el menor

$$\begin{vmatrix} 1 & 1 \\ -3 & 4 \end{vmatrix} \neq 0$$

Las filas básicas son ahora $\{1, 2\}$, y las columnas básicas $\{2, 4\}$. A continuación orlamos con la tercera fila y la quinta columna (es inmediato ver que, orlando

con la primera o la tercera columna no podemos conseguir un menor distinto de cero). El menor

$$\begin{vmatrix} 1 & 1 & 3 \\ -3 & 4 & 1 \\ 2 & -5 & -4 \end{vmatrix} = 0$$

De este modo, orlando con la tercera fila no hay forma de obtener un menor distinto de cero, por lo que podemos eliminar dicha fila.

Orlamos con la cuarta fila y la columna primera; el menor

$$\begin{vmatrix} 0 & 1 & 1 \\ 0 & -3 & 4 \\ 1 & 1 & 1 \end{vmatrix} \neq 0$$

Luego las filas básicas son ahora $\{1, 2, 4\}$ y las columnas básicas $\{1, 2, 4\}$.

Orlamos con la quinta fila y la tercera columna,

$$\begin{vmatrix} 0 & 1 & 0 & 1 \\ 0 & -3 & 0 & 4 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{vmatrix} = 0$$

Por tanto podemos eliminar la columna tercera. Finalmente, orlando con la quinta columna:

$$\begin{vmatrix} 0 & 1 & 1 & 3 \\ 0 & -3 & 4 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{vmatrix} \neq 0$$

Puesto que no quedan más filas por orlar, el $\text{rango}(A) = 4$.

Es importante resaltar que, en todo el procedimiento, una vez que conseguimos un menor básico de un cierto orden, no lo perdemos, es decir, a la hora de buscar menores básicos de orden superior usamos siempre el menor que ya tenemos completándolo con más filas y columnas, esto es, *orlándolo*.

### Cálculo del rango mediante el método de Gauss

Existe una alternativa para el cálculo del rango de una matriz usando el método de Gauss descrito en la sección 2.3. Dado que las operaciones involu-

cradas en el método de Gauss no modifican el rango (como consecuencia de las propiedades de los determinantes), para realizar este cálculo procedemos a triangular la matriz del mismo modo que hacemos con un sistema, es decir, haciendo nulos todos los elementos que quedan por debajo de la diagonal principal. El rango de la matriz resultará igual al número de filas no nulas que resulten tras la triangulación.

**Ejemplo 2.16**

Calcular el rango de la matriz

$$A = \begin{pmatrix} 1 & 6 & 11 & 16 \\ 2 & 7 & 12 & 17 \\ 3 & 8 & 13 & 18 \\ 4 & 9 & 14 & 19 \\ 5 & 10 & 15 & 20 \end{pmatrix} \in \mathcal{M}_{5\times 4}(\mathbb{R})$$

Triangulamos mediante el método de Gauss, haciendo ceros en la primera columna por debajo de la primera fila, y luego haciendo ceros en la segunda columna, por debajo de la segunda fila

$$\begin{pmatrix} 1 & 6 & 11 & 16 \\ 2 & 7 & 12 & 17 \\ 3 & 8 & 13 & 18 \\ 4 & 9 & 14 & 19 \\ 5 & 10 & 15 & 20 \end{pmatrix} \xrightarrow[F_5-5F_1]{\substack{F_2-2F_1 \\ F_3-3F_1 \\ F_4-4F_1}} \begin{pmatrix} 1 & 6 & 11 & 16 \\ 0 & -5 & -10 & -15 \\ 0 & -10 & -20 & -30 \\ 0 & -15 & -30 & -45 \\ 0 & -20 & -40 & -60 \end{pmatrix}$$

$$\xrightarrow[F_5-4F_2]{\substack{F_3-2F_2 \\ F_4-3F_2}} \begin{pmatrix} 1 & 6 & 11 & 16 \\ 0 & -5 & -10 & -15 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Dado que la matriz obtenida es triangular y que el número de filas no nulas es dos, el rango de la matriz $A$ es 2.

2 7

## APLICACIÓN DE LOS DETERMINANTES AL CÁLCULO DE LA INVERSA

Finalmente, veamos cómo se usan los determinantes para calcular la inversa de una matriz.

**Definición 2.17**

Dada una matriz cuadrada $A$, se denomina *adjunta* de una matriz (o *matriz de cofactores*) de $A$ a la matriz $\text{adj}(A) \in \mathcal{M}_n(\mathbb{K})$ dada por

$$(\text{adj}\, A)_{ij} = (-1)^{i+j}|A_{ij}|$$

Es decir, la adjunta de una matriz es una nueva matriz en la que cada elemento $ij$ es igual al determinante de la matriz adjunta de ese elemento, multiplicada por $(-1)^{i+j}$, es decir, con un determinado signo, que dependerá de la "posición" de ese elemento. A partir de la matriz de cofactores se obtiene la inversa de una matriz, según muestra el siguiente resultado.

**Teorema 2.5 (Inversa de una matriz)**

Sea $A \in \mathcal{M}_n(\mathbb{K})$ una matriz invertible. Entonces

$$A^{-1} = \frac{1}{\det A}\,\text{adj}(A)^T$$

*Demostración:*
Denotemos por $c_{ij} = (\text{adj}(A))_{ij}$. Entonces,

$$A\,\text{adj}(A)^T = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{nn} \end{pmatrix}$$

$$= \begin{cases} (A\,\text{adj}(A)^T)_{jj} = \sum_{k=1}^{n} a_{jk}c_{jk} & \text{si } j = i \\ (A\,\text{adj}(A)^T)_{ij} = \sum_{k=1}^{n} a_{ik}c_{jk} & \text{si } j \neq i \end{cases}$$

Ahora bien, por la definición de la adjunta y el Teorema 2.3

$$(A\,\mathrm{adj}(A)^T)_{jj} = \sum_{k=1}^{n} a_{jk}c_{jk} = \sum_{k=1}^{n}(-1)^{j+k}a_{jk}|A_{jk}| = |A| \qquad (2.4)$$

Por otra parte, si $B$ es la matriz siguiente, en la que las filas $i$ y $j$ son iguales,

$$B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{matrix} \\ \\ \longleftarrow \text{Fila } i \\ \\ \longleftarrow \text{Fila } j \\ \\ \\ \end{matrix}$$

y desarrollamos su determinante por la fila $j$, entonces

$$|B| = \sum_{k=1}^{n} a_{ik}c_{jk} = 0 \qquad (2.5)$$

pues $B$ tiene dos filas iguales. De (2.4) y (2.5) se sigue el resultado.

■

**Nota 2.5**

Obsérvese que el resultado anterior solo tiene sentido si $|A| \neq 0$, que es, precisamente, la condición de no singularidad obtenida en la Proposición 2.15.

**Ejemplo 2.17**

Calcular la inversa de la matriz

$$A = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 1 & 2 \\ 0 & 1 & 3 \end{pmatrix}$$

En primer lugar calculamos $|A| = 7$. Puesto que éste no es cero, procedemos a

calcular la matriz adjunta,

$$\operatorname{adj}(A) = \begin{pmatrix} 1 & 3 & -1 \\ -6 & 3 & -1 \\ 4 & -3 & 3 \end{pmatrix}$$

y finalmente,

$$A^{-1} = \frac{1}{7}\begin{pmatrix} 1 & -6 & 4 \\ 3 & 3 & -3 \\ -1 & -1 & 3 \end{pmatrix}$$

2 8

## CÁLCULO CON PYTHON

En este tema hemos introducido las matrices, con las que vamos a realizar multitud de cálculos a lo largo de todo el texto, y en esta sección nos centraremos en mostrar cómo se pueden llevar a cabo con Python. No obstante, insistimos encarecidamente en la importancia que tiene saber hacer realizar todas estas operaciones "a mano", sin necesidad de usar el ordenador.

En el apéndice B introdujimos los módulos NumPy y SymPy, gracias a los cuales es posible trabajar fácilmente con matrices:

```
>>> from numpy import matrix
>>> a=matrix('1. 2 ; 3 4')
>>> a
matrix([[ 1.,  2.],
        [ 3.,  4.]])
```

Las operaciones más comunes se realizan de forma natural: por ejemplo:

$$3\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 9 & 12 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 8 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 13 & 22 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^2 = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix}$$

se pueden realizar con el siguiente código:

```
>>> b=matrix('3. 2; 1 4')
```

```
7  >>> 3*a
8  matrix([[  3.,   6.],
9          [  9.,  12.]])
10 >>> a+b
11 matrix([[ 4.,  4.],
12         [ 4.,  8.]])
13 >>> a*b
14 matrix([[  5.,  10.],
15         [ 13.,  22.]])
16 >>> a**2
17 matrix([[  7.,  10.],
18         [ 15.,  22.]])
```

También tenemos atributos sencillos con los que construir la matriz traspuesta y la inversa:

```
19 >>> a.T
20 matrix([[  1.,   3.],
21         [  2.,   4.]])
22 >>> a.I
23 matrix([[-2. ,  1. ],
24         [ 1.5, -0.5]])
```

es decir,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

Recordemos que el módulo NumPy trabaja con números reales, por lo que en determinados momentos nos encontraremos con resultados como estos:

```
25 >>> c=matrix('1 1; 4 1')
26 >>> c.I
27 matrix([[-0.33333333,  0.33333333],
28         [ 1.33333333, -0.33333333]])
```

Si queremos cálculos exactos, debemos usar el módulo SymPy:

```
29 >>> from sympy import Matrix
30 >>> C=Matrix(c)
31 >>> C.inv()
32 [-1/3,  1/3]
33 [ 4/3, -1/3]
```

Nótese que la función para definir matrices en SymPy no es la misma que en NumPy (Python es sensible a las mayúsculas). En la línea 30, usamos la función `Matrix` para convertir una matriz de NumPy en una de SymPy. El cálculo de la matriz inversa ahora no es posible con el atributo `I`, pues no está definido para objetos de SymPy. Es por eso que debemos usar el método `inv()` (línea 31).

El orden de la matriz se obtiene a través del atributo `shape`, y el acceso a las filas o columnas de la matriz se obtiene rápidamente mediante *slicing*, recordando que los índices en Python comienzan en 0:

```
>>> a.shape
(2, 2)
>>> a[0,:]
matrix([[ 1.,  2.]])
>>> a[:,1]
matrix([[ 2.],
        [ 4.]])
```

El operador de *slicing* proporciona el entorno adecuado para realizar las operaciones involucradas en el método de Gauss: el siguiente código

```
>>> a=matrix('1 2 0 1; 1 1 2 0; 1 0 2 1')
>>> A=Matrix(a)
>>> A[1,:]=A[1,:]-A[0,:]
>>> A[2,:]=A[2,:]-A[0,:]
>>> A
[1,  2, 0,  1]
[0, -1, 2, -1]
[0, -2, 2,  0]
>>> A[2,:]=A[2,:]-2*A[1,:]
>>> A
[1,  2,  0,  1]
[0, -1,  2, -1]
[0,  0, -2,  2]
```

corresponde a

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 0 & 2 & 1 \end{pmatrix} \xrightarrow[F_3-F_1]{F_2-F_1} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & -1 & 2 & -1 \\ 0 & -2 & 2 & 0 \end{pmatrix}$$

$$\xrightarrow{F_3-2F_2} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -2 & 2 \end{pmatrix}$$

Obsérvese que hemos trabajado con el objeto de SymPy para obtener resultados exactos, aunque también se podrían haber realizado los cálculos con el objeto de Numpy.

El *slicing* también es útil para extraer submatrices de una matriz:

```
>>> A
[1, 2, 0, 1]
[1, 1, 2, 0]
```

```
[1, 0, 2, 1]
>>> B=A[:,0:3]
>>> B
[1, 2, 0]
[1, 1, 2]
[1, 0, 2]
>>> C=A[:,3]
>>> C
[1]
[0]
[1]
```

Prestar atención al manejo de los índices en Python. La operación `A[:,0:3]` extrae las tres primeras columnas de `A` (es decir, hace variar el segundo índice entre 0 y 3, sin llegar a éste último).

Los módulos NumPy y SymPy también permiten calcular el determinante de matrices cuadradas, pero cada uno lo hace de forma distinta:

```
>>> from numpy import linalg,matrix
>>> from sympy import Matrix
>>> a=matrix('1 3 5; 2 3 1; -1 1 1')
>>> b=Matrix([[2,5,1],[2,1,-1],[0,-2,2]])
>>> linalg.det(a)
18.0
>>> b.det()
-24
```

$$\begin{vmatrix} 1 & 3 & 5 \\ 2 & 3 & 1 \\ -1 & 1 & 1 \end{vmatrix} = 18 \qquad \begin{vmatrix} 2 & 5 & 1 \\ 2 & 1 & -1 \\ 0 & -2 & 2 \end{vmatrix} = -24$$

Lamentablemente, ninguno de los dos módulos posee una función para calcular el rango de una matriz de forma directa, aunque se puede obtener por otros medios. En concreto, en SymPy encontramos el método `rref()` que nos proporciona el rango de una matriz de forma indirecta. Este método equivale a realizar operaciones elementales en la matriz similares a las hechas en el ejemplo 2.10 (es decir, tratando de "diagonalizar" la matriz mediante el método de Gauss). Por ejemplo,

```
>>> a=Matrix([[1,6,11,16],[2,7,12,17],
[3,8,13,18],[4,9,14,19],[5,10,15,20]])
>>> a.rref()
([1, 0, -1, -2]
[0, 1,  2,  3]
[0, 0,  0,  0]
[0, 0,  0,  0]
```

```
[0,  0,   0,   0], [0,  1])
```

equivale a

$$\begin{pmatrix} 1 & 6 & 11 & 16 \\ 2 & 7 & 12 & 17 \\ 3 & 8 & 13 & 18 \\ 4 & 9 & 14 & 19 \\ 5 & 10 & 15 & 20 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Observamos que el resultado de esta operación nos da una matriz que primero ha sido triangulada mediante el método de Gauss (como en el ejemplo 2.16), y posteriormente se han calculado ceros sobre la diagonal principal y se han hecho unos en la diagonal, allí donde esto ha sido posible (es decir, algo similar al método de Gauss-Jacobi). Dado que esta matriz tiene dos filas no nulas, su rango es dos.

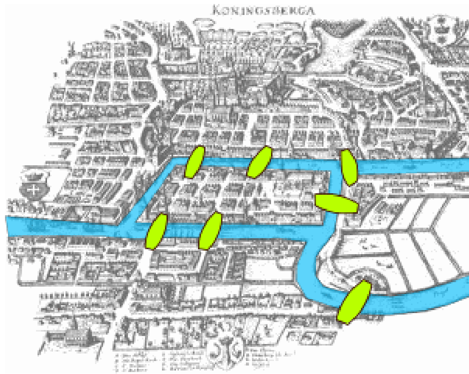
## 2 9 BREVE INTRODUCCIÓN A LA TEORÍA DE GRAFOS

La teoría de grafos es una rama de las Matemáticas que ha recibido gran atención debido a sus múltiples aplicaciones en campos tan variados como la Ingeniería, la Física, la Psicología o la Sociología, entre otros. Tiene su origen en el problema de los siete puentes de Königsberg,[10] que fue resuelto por Euler en 1736, y que consistía en averiguar si es posible dar un paseo por todos los puentes que conectan la ciudad con la isla que forma el río Pregolya al atravesarla, de manera que se cruce cada puente una sola vez y se vuelva al punto de partida (véase la figura 2.4a).

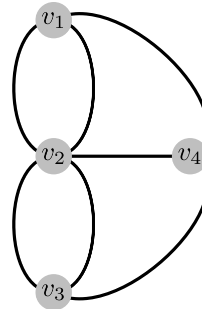
Este problema admite una representación esquemática en forma de *grafo*. Un grafo es un objeto formado por un conjunto de puntos denominados *vértices* y un conjunto de pares de vértices llamados *aristas*. Esquemáticamente un grafo se representa por una gráfica como la de la figura 2.4b, que corresponde al esquema de conexiones en el problema de los puentes de Königsberg. En este ejemplo, cada vértice representa cada una de las zonas terrestres que quedan divididas por el río, y las aristas corresponden a los puentes que las conectan. La información que nos proporciona el gráfico es la de las conexiones entre sus vértices: por ejemplo, el vértice $v_1$ está conectado con el $v_2$ mediante dos aristas, y con el $v_4$ sólo mediante una. Como podemos apreciar, en lo que se refiere al problema, la información del grafo es equivalente a la del mapa que muestra la situación de los puentes.

Podemos representar la información que proporciona este grafo mediante una matriz, denominada *matriz de adyacencia* que nos da el número de conexiones en el grafo, de tal modo que el elemento $a_{ij}$ de la matriz es igual al número

[10] Actual Kaliningrado.




(a) Mapa de Königsberg en la época de Euler con la situación de los puentes




(b) Grafo que representa el problema

Figura 2.4: El problema de los siete puentes de Königsberg

de aristas que conecta el vértice $i$ con el vértice $j$. De este modo, la matriz del grafo de la figura 2.4b es

$$\begin{pmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \tag{2.6}$$

Las sucesivas potencias de la matriz de adyacencia de un grafo tienen un significado peculiar. Por ejemplo, para la matriz $A$ de (2.6) se tiene que

$$B = A^2 = \begin{pmatrix} 5 & 1 & 5 & 2 \\ 1 & 9 & 1 & 4 \\ 5 & 1 & 5 & 2 \\ 2 & 4 & 2 & 3 \end{pmatrix}$$

¿Qué significa que el elemento $b_{12} = 1$, o que $b_{14} = 4$? En este caso, el hecho de que $b_{12} = 1$ significa que hay un único camino de longitud 2 (es decir, que se ha de recorrer en dos etapas) que une el vértice $v_1$ con $v_2$ (el que une $v_1 \to v_4 \to v_2$, véase la figura 2.5); es lo que se denomina un 2-*camino*. Que $b_{24} = 4$ significa que podemos unir los vértices $v_1$ y $v_4$ mediante 2-caminos de cuatro modos diferentes. ¿Puede el lector encontrarlos?

En definitiva, la entrada $b_{ij}$ de la matriz $B = A^2$ nos da el número de 2-caminos que unen los vértices $v_i$ con $v_j$. El significado de los sucesivas potencias

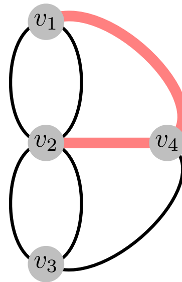



Figura 2.5: Camino $v_1 \to v_4 \to v_2$

de $A$ es análogo: las entradas de la potencia $n$ de $A$ nos proporcionan el número de $n$-caminos que unen los correspondientes vértices.

Los grafos también pueden ser *dirigidos*. Por ejemplo, supongamos que tenemos una serie de lugares (vértices) unidos por caminos (aristas) pero que alguno de éstos solo puede ser recorrido en una única dirección (véase la figura 2.6a). La matriz de adyacencia en este caso sería

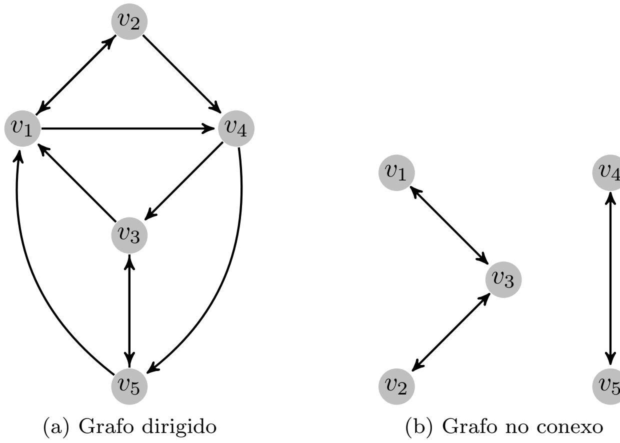
$$C = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

que, a diferencia del ejemplo anterior, no es simétrica

La potencia $n$ de esta matriz también nos proporciona el número de $n$-caminos que conectan el vértice $i$ con el $j$, pero en este caso, respetando la direccionalidad. Por ejemplo,

$$C^2 = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Nótese que hay dos 2-caminos que unen el vértice $v_4$ con $v_1$, pero solo un 2-camino que une $v_1$ con $v_4$.




(a) Grafo dirigido (b) Grafo no conexo

Figura 2.6: Grafos

En todo grafo de $k$ vértices, está claro que cualquier trayectoria que una un vértice $v_i$ con $v_j$ tendrá como máximo longitud $k-1$, de manera que si $A$ es su matriz de adyacencia se verifica que la matriz

$$A + A^2 + \cdots + A^{k-1}$$

tiene un cero en la posición $ij$ si y solo si no existe ninguna trayectoria (de cualquier longitud) que una el vértice $v_i$ con el $v_j$.

Por ejemplo, la matriz de adyacencia del grafo de la figura 2.6b es

$$D = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Es evidente que en ese grafo no hay forma de llegar desde los vértices $v_1$, $v_2$ ó $v_3$ hasta los vértices $v_4$ ó $v_5$, hecho que se puede comprobar calculando las posiciones de los ceros de la matriz

$$D + D^2 + D^3 + D^4 = \begin{pmatrix} 3 & 3 & 3 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 3 & 3 & 6 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 \end{pmatrix}$$

Estamos ante lo que se denomina un grafo *no conexo*,[11] pues hay pares de vértices que no se pueden unir mediante ninguna trayectoria, lo que queda reflejado por la matriz anterior.

2 10

## EJERCICIOS

### Ejercicios de repaso

**E.1** Los precios en euros de las entradas a un parque temático para adultos (AD) y niños y jubilados (NJ) en temporada alta (TA), temporada media (TM) y temporada baja (TB) vienen dados por la matriz $P$. El número de asistentes (en miles) a dicho parque a lo largo de un año viene dado por la matriz $N$:

$$P = \begin{array}{c} \\ \text{AD} \\ \text{NJ} \end{array}\begin{array}{c} \begin{array}{ccc} \text{TA} & \text{TM} & \text{TB} \end{array} \\ \begin{pmatrix} 25 & 20 & 14 \\ 20 & 15 & 7 \end{pmatrix} \end{array} \qquad N = \begin{array}{c} \\ \text{TA} \\ \text{TM} \\ \text{TB} \end{array}\begin{array}{c} \begin{array}{cc} \text{AD} & \text{NJ} \end{array} \\ \begin{pmatrix} 500 & 600 \\ 350 & 300 \\ 125 & 100 \end{pmatrix} \end{array}$$

Se pide:

(a) Obtener, si es posible, las matrices: $R_1 = NP$ y $R_2 = PN$.

(b) ¿A cuántos euros asciende la recaudación total correspondiente a los niños y jubilados? ¿Y la correspondiente a temporada baja?

(c) ¿Qué elemento de $R_1$ o $R_2$ nos proporciona información sobre la recaudación total correspondiente a los adultos?

(d) ¿A cuántos euros asciende la recaudación total?

**E.2** Si $A = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ y $B = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, calcular $A^T B$, $B^T A$, $AB^T$ y $BA^T$.

**E.3** Realizar las siguientes operaciones:

$$A + 2B; \quad iA - 5B; \quad AB; \quad BA; \quad A\overline{B};$$

donde

$$A = \begin{pmatrix} 3 & i & 5-i \\ 1-i & -i & 1+2i \\ 1 & 0 & -3+i \end{pmatrix}, \qquad B = \begin{pmatrix} -1-i & 1+i & 3i \\ 3+i & -i & 1 \\ 3 & 5-2i & -2i \end{pmatrix}$$

[11] Intuitivamente, el concepto de conexidad significa que el grafo puede ser separado en subgrafos que no están conectados por ninguna arista.

E.4 Escribir los siguientes sistemas en forma matricial y resolverlos mediante el método de Gauss:

$$\text{(a)}\quad \left.\begin{array}{rcl} 2x+3y+z &=& -1\\ 3x+3y+z &=& 1\\ 2x+4y+z &=& -2 \end{array}\right\} \qquad \text{(b)}\quad \left.\begin{array}{rcl} x_1-2x_2-x_3 &=& 0\\ 3x_1-5x_2-2x_3 &=& 5\\ 3x_1+x_2-2x_3 &=& 2 \end{array}\right\}$$

E.5 Encontrar la inversa de las siguientes matrices usando el método de Gauss:

$$\text{(a)}\quad \begin{pmatrix} 1 & 0 & 0\\ 2 & 1 & -1\\ 3 & 1 & 4 \end{pmatrix} \qquad \text{(b)}\quad \begin{pmatrix} 1 & 2 & 3 & 4\\ 1 & 3 & 2 & 6\\ -2 & -2 & -6 & 0\\ 1 & 4 & 1 & 7 \end{pmatrix}$$

E.6 Escribir las matrices del ejercicio 5 como producto de matrices elementales.

E.7 Calcular el rango de las siguientes matrices:

$$\text{(a)}\quad \begin{pmatrix} 1 & 3 & 7 & 4\\ 6 & -5 & -1 & 2\\ 7 & -2 & 6 & 1 \end{pmatrix} \qquad \text{(b)}\quad \begin{pmatrix} 2 & 3 & 1 & -4 & 2\\ 1 & 2 & 1 & -2 & 3\\ 2 & 3 & 0 & 2 & 1\\ 1 & 2 & 0 & 2 & 1 \end{pmatrix}$$

E.8 Demostrar, sin necesidad de calcularlos, que los siguientes determinantes son nulos:

$$\text{(a)}\quad \begin{vmatrix} 1 & 2 & 3 & 4 & 5\\ 2 & 3 & 4 & 5 & 6\\ 3 & 4 & 5 & 6 & 7\\ 4 & 5 & 6 & 7 & 8\\ 5 & 6 & 7 & 8 & 9 \end{vmatrix} \qquad \text{(b)}\quad \begin{vmatrix} 1 & 2^2 & 3^2 & 4^2\\ 2^2 & 3^2 & 4^2 & 5^2\\ 3^2 & 4^2 & 5^2 & 6^2\\ 4^2 & 5^2 & 6^2 & 7^2 \end{vmatrix}$$

### Problemas

E.9 Hallar la matriz cuadrada de segundo orden que expresa la suma $X^2+Y^2$, siendo $X$ e $Y$ dos matrices, también de segundo orden, que satisfacen el sistema:

$$\left.\begin{array}{l} 5X+3Y=A\\ 3X+2Y=B \end{array}\right\}$$

donde $A=\begin{pmatrix} 2 & 0\\ -4 & 15 \end{pmatrix}$ y $B=\begin{pmatrix} 1 & -1\\ -2 & 9 \end{pmatrix}$.

E.10 Calcular todas las matrices de orden dos tales que:

(a) su cuadrado sea la matriz nula;

(b) sean idempotentes, es decir, $A^2 = A$.

**E.11** Hallar todas las matrices $B \in \mathcal{M}_2(\mathbb{R})$ que conmutan con la matriz

$$A = \begin{pmatrix} 1 & 3 \\ -5 & 2 \end{pmatrix}$$

**E.12** Encontrar la inversa de las siguientes matrices usando el método de Gauss:

$$\text{(a)} \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix} \qquad \text{(b)} \begin{vmatrix} 1 & -2 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

***E.13** Sabiendo que los números 58786, 30628, 80743, 12831 y 16016, son múltiplos de 13, probar que el siguiente determinante también lo es:

$$\begin{vmatrix} 5 & 8 & 7 & 8 & 6 \\ 3 & 0 & 6 & 2 & 8 \\ 1 & 2 & 8 & 3 & 1 \\ 8 & 0 & 7 & 4 & 3 \\ 1 & 6 & 0 & 1 & 6 \end{vmatrix}$$

***E.14** Demostrar que el conocido como *determinante de Vandermonde*[12] tiene por valor:

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i)$$

Sugerencia: aplicar el principio de inducción y realizar las siguientes operaciones: $F_{j+1} - x_{n+1}F_j$, para $j = 1, \dots, n$.

[12]Denominado así por el matemático francés Alexandre Vandermonde. Aunque hizo interesantes contribuciones al estudio de los determinantes, éste en concreto no aparece en sus trabajos. Se cree que la atribución es debida a cierta confusión con la notación que empleaba.

**E.15** Evaluar los siguientes determinantes:

$$\text{(a)} \begin{vmatrix} 1 & 2 & 3 & \cdots & n \\ 1 & x+1 & 3 & \cdots & n \\ 1 & 2 & x+1 & \cdots & n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & x+1 \end{vmatrix}$$

$$\text{(b)} \begin{vmatrix} 1 & 2 & 3 & \cdots & n-1 & n \\ 1 & 3 & 3 & \cdots & n-1 & n \\ 1 & 2 & 5 & \cdots & n-1 & n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & 2n-3 & n \\ 1 & 2 & 3 & \cdots & n-1 & 2n-1 \end{vmatrix} \qquad \text{(c)} \begin{vmatrix} 1 & n & n & \cdots & n \\ n & 2 & n & \cdots & n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & n & n & \cdots & n \end{vmatrix}$$

***E.16** Probar que:

$$\text{(a)} \begin{vmatrix} x+a_1 & a_2 & a_3 & \cdots & a_n \\ a_1 & x+a_2 & a_3 & \cdots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & x+a_n \end{vmatrix} = x^{n-1}\left(x + \sum_{i=1}^n a_i\right)$$

$$\text{(b)} \begin{vmatrix} 2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{vmatrix} = n+1$$

**E.17** Encontrar el rango de la matrices

$$\text{(a)} \begin{pmatrix} 1 & a & -1 & 2 \\ 2 & -1 & a & 5 \\ 1 & 10 & -6 & 1 \end{pmatrix} \qquad \text{(b)} \begin{pmatrix} a & -1 & 0 \\ 1 & a^2 & 1+a^2 \\ 2 & 2 & 2+2a \end{pmatrix}$$

según los valores de $a$.

**E.18** Halla el rango de la matriz

$$\begin{pmatrix} 1 & 1 & 1 & a \\ 1 & 1+a & 1 & 2a \\ b & 1 & 1+a & 0 \end{pmatrix}$$

en función de los valores $a$ y $b$.

*E.19 Hallar la inversa de la siguiente matriz:

$$\begin{vmatrix} 1+x & 1 & 1 & \cdots & 1 \\ 1 & 1+x & 1 & \cdots & 1 \\ 1 & 1 & 1+x & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1+x \end{vmatrix}$$

### Ejercicios teóricos

E.20 Sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$, y sean $F_i \in \mathcal{M}_{1\times m}$ y $C_j \in \mathcal{M}_{n\times 1}$ las matrices fila y columna respectivamente, cuyos elementos son todos nulos salvo un 1 en la posición $1i$ de $F_i$ y $j1$ de $C_j$. Calcular los productos $F_iA$ y $AC_j$.

E.21 Sean $A, B \in \mathcal{M}_n(\mathbb{K})$ dos matrices simétricas. Probar que $AB$ es simétrica si y solo si $A$ y $B$ conmutan.

E.22 Probar que si $A \in \mathcal{M}_n(\mathbb{K})$ es una matriz simétrica, entonces $A^k$ es simétrica para todo $k \in \mathbb{N}$.

E.23 Si $A$ es una matriz regular, probar que $A$ es simétrica si y solo si $A^{-1}$ es simétrica.

E.24 Se dice que $A \in \mathcal{M}_n(\mathbb{K})$ es antisimétrica si $A = -A^T$. Probar que dada cualquier matriz cuadrada $M \in \mathcal{M}_n(\mathbb{K})$, $M - M^T$ es una matriz antisimétrica. Usar este hecho para demostrar que cualquier matriz cuadrada se puede escribir como suma de una matriz simétrica y otra antisimétrica.

E.25 Sean $A \in \mathcal{M}_n(\mathbb{K})$ una matriz regular tal que $AB = 0$, donde $B \in \mathcal{M}_{n\times p}(\mathbb{K})$. Probar que $B = 0$.

E.26 Una matriz $A \in \mathcal{M}_n(\mathbb{K})$ se dice nilpotente si existe $k \in \mathbb{N}$ tal que $A^k = 0$. Probar que si $A$ es nilpotente, entonces $I - A$ es invertible y su inversa es

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^{k-1}$$

E.27 Probar la Proposición 2.17.

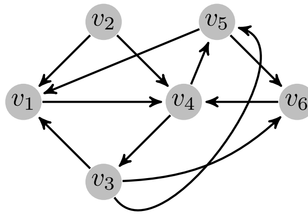
### Ejercicios adicionales

E.28 Dada la matriz

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

dibujar el grafo que representa a dicha matriz y averiguar, usando Python, el número de 3-caminos que conectan los vértices $v_1$ y $v_3$ del mismo.

E.29 Dado el grafo




escribir su matriz de adyacencia y usar Python para averiguar si existe algún vértice no alcanzable.

# 3 Sistemas de ecuaciones lineales

En el tema anterior se introdujeron los sistemas de ecuaciones lineales como una de las aplicaciones habituales de las matrices y en éste vamos a recordar resultados, como la Regla de Cramer o el Teorema de Rouché-Frobenius, que el lector habrá tenido la oportunidad de ver en cursos anteriores. También introduciremos algunos aspectos nuevos relacionados con el Álgebra Lineal Numérica, estudiando algunos métodos computacionales diseñados para resolver sistemas de ecuaciones. Veremos cómo usar Python tanto en la resolución de sistemas como en la programación de métodos numéricos y finalizaremos mostrando una aplicación del uso de sistemas lineales en la resolución de ecuaciones diferenciales ordinarias.

## 3 1 SISTEMAS DE ECUACIONES LINEALES: PRIMERAS DEFINICIONES

Recordemos que un sistema de ecuaciones lineal es una expresión de la forma

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}\right\} \tag{3.1}$$

en el que los elementos que intervienen ya aparecieron en la Definición 2.9. El concepto de solución dado en la Definición 2.10, junto con la definición de *compatibilidad* que presentamos a continuación, conforman prácticamente todos los elementos necesarios para trabajar con sistemas de ecuaciones lineales.

**Definición 3.1**

Un sistema se dice *compatible* si posee solución, e *incompatible* si no posee solución. Un sistema compatible se dice *determinado* si su solución es única. En caso contrario se dirá *indeterminado*.

La Proposición 2.7 permite poner en marcha el método de Gauss, que ya hemos visto en el tema anterior (pág. 49), y que recordamos con nuevos ejemplos:

**Ejemplo 3.1**

Resolver mediante el método de Gauss los siguientes sistemas:

(I)

$$\left.\begin{array}{rcl} x_1 - x_2 + x_3 & = & 1 \\ x_1 + x_2 + x_3 & = & 0 \\ 2x_1 - 2x_2 + 2x_3 & = & 3 \end{array}\right\}$$

Recordemos que el método de Gauss consiste en triangular el sistema mediante transformaciones entre sus ecuaciones que lleven a un sistema equivalente que se resuelve mediante una sencillo proceso de subida. Reescribiendo el sistema en forma matricial, las operaciones conducentes a la triangulación son, en este caso,

$$\left(\begin{array}{rrr|r} 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 2 & -2 & 2 & 3 \end{array}\right) \xrightarrow[F_3-2F_1]{F_2-F_1} \left(\begin{array}{rrr|r} 1 & -1 & 1 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

Si ahora reescribimos el sistema resulta que la última ecuación se escribe $0 = 1$!!! Se trata por tanto de un sistema incompatible.

(II)

$$\left.\begin{array}{rcl} x_1 + 3x_2 - x_3 + x_4 & = & 1 \\ -2x_1 + x_2 + 2x_3 & = & 7 \\ x_2 - x_4 & = & 0 \\ 2x_1 + 6x_2 - 2x_3 + 2x_4 & = & 2 \end{array}\right\}$$

El método de Gauss resulta ahora:

$$\left(\begin{array}{rrrr|r} 1 & 3 & -1 & 1 & 1 \\ -2 & 1 & 2 & 0 & 7 \\ 0 & 1 & 0 & -1 & 0 \\ 2 & 6 & -2 & 2 & 2 \end{array}\right) \xrightarrow[F_4-2F_1]{F_2+2F_1} \left(\begin{array}{rrrr|r} 1 & 3 & -1 & 1 & 1 \\ 0 & 7 & 0 & 2 & 9 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Podemos ver que la última ecuación es $0 = 0$, con lo que no aporta información alguna al sistema y podemos eliminarla. Finalmente, intercambiando

las filas segunda y tercera para simplificar los cálculos, se tiene

$$\left(\begin{array}{cccc|c} 1 & 3 & -1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 7 & 0 & 2 & 9 \end{array}\right) \xrightarrow{F_3-7F_2} \left(\begin{array}{cccc|c} 1 & 3 & -1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 9 & 9 \end{array}\right)$$

con lo que el sistema queda triangulado. En este caso, al realizar el proceso de subida obtenemos $x_4 = 1$, $x_2 = 1$, sin embargo la primera ecuación resulta $x_1 = x_3 - 3$. Puesto que tenemos más incógnitas que ecuaciones, $x_3$ queda libre para tomar el valor que uno desee, obteniéndose así un valor para $x_1$. Puesto que $x_3$ puede tomar infinitos valores, estamos ante un sistema compatible indeterminado. Un conjunto de soluciones puede expresarse por $(\alpha-3, 1, \alpha, 1)$ con $\alpha \in \mathbb{R}$. Como sólo tenemos una incógnita libre (a la que se denomina *parámetro*) se dice que el sistema tiene un *grado de libertad*. Nótese que el número de grados de libertad es la diferencia entre incógnitas y número de ecuaciones relevantes, esto es, ecuaciones que no resultan eliminadas en el transcurso de la aplicación del método de Gauss.

(iii)

$$\left.\begin{array}{rcr} x_1 - x_2 + 2x_3 - x_4 & = & 1 \\ -x_1 + 2x_2 - x_3 - x_4 & = & -2 \\ 3x_1 - 4x_2 + 5x_3 - x_4 & = & 4 \end{array}\right\}$$

Observemos que este sistema no puede ser determinado, puesto que tenemos más incógnitas que ecuaciones. Procediendo como antes,

$$\left(\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & 1 \\ -1 & 2 & -1 & -1 & -2 \\ 3 & -4 & 5 & -1 & 4 \end{array}\right) \xrightarrow[F_3-3F_1]{F_2+F_1} \left(\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & 1 \\ 0 & 1 & 1 & -2 & -1 \\ 0 & 1 & -2 & 1 & 1 \end{array}\right)$$

Si a la fila tres le sumamos la fila dos (obsérvese que tienen signos opuestos) la tercera ecuación resulta irrelevante, de manera que llegamos a un sistema triangular de dos ecuaciones con cuatro incógnitas, tratándose por tanto de un sistema compatible indeterminado. En este caso tenemos dos grados de libertad, lo cual significa que las soluciones del sistema quedan determinadas escogiendo dos parámetros; en este caso dejamos libres $x_3$ y $x_4$. La solución general puede expresarse como

$$(-3\alpha + 3\beta, -1 - \alpha + 2\beta, \alpha, \beta), \quad \alpha, \beta \in \mathbb{R}$$

En definitiva, hemos visto a través de ejemplos cómo el método de Gauss es capar de reflejar el carácter de un sistema, es decir, si es o no compatible, y el número de grados de libertad que posee.

A continuación vamos a centrarnos en un tipo de sistemas que jugará un papel muy relevante a lo largo del texto.

### Sistemas Homogéneos

**Definición 3.2**

Un sistema de la forma (3.1) se dirá *homogéneo*, si el término independiente $b_i = 0$, para todo $i = 1, \dots, m$.

Es inmediato comprobar que todo sistema homogéneo es compatible, pues la $n$-upla $(0, \dots, 0)$ es solución del mismo (comúnmente denominada solución *trivial*). Por otra parte, la aplicación del método de Gauss a un sistema homogéneo basta realizarla sobre la matriz de los coeficientes, puesto que la columna correspondiente al término independiente se mantiene siempre nula. Si después de la triangulación por Gauss y la posterior eliminación de filas nulas la matriz resultante es cuadrada (es decir, el número de grados de libertad del sistema es cero), el sistema es compatible determinado, y su única solución es la trivial. En otro caso, el sistema será compatible indeterminado.

**Proposición 3.1**

Dado el sistema homogéneo

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n & = & 0 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n & = & 0 \end{array}\right\} \qquad (3.2)$$

si $(\bar{x}_1, \dots, \bar{x}_n)$ es solución, entonces $(\alpha\bar{x}_1, \dots, \alpha\bar{x}_n)$ también es solución del sistema, $\forall \alpha \in \mathbb{R}$.

Del mismo modo, si $(\bar{x}_1, \dots, \bar{x}_n)$ e $(\bar{y}_1, \dots, \bar{y}_n)$ son soluciones de (3.2), entonces la $n$-upla $(\bar{x}_1 + \bar{y}_1, \dots, \bar{x}_n + \bar{y}_n)$ también es solución de (3.2).

La demostración de este resultado es trivial y se deja como ejercicio al lector. Obsérvese que la tesis del resultado es cierta con independencia de que el sistema sea o no determinado.

Las soluciones de un sistema homogéneo son relevantes en relación con las soluciones de uno no homogéneo, como pone de manifiesto el siguiente resultado.

**Proposición 3.2**

Si $(\bar{x}_1, \ldots, \bar{x}_n)$ es solución del sistema (3.1), entonces todas sus restantes soluciones se escriben de la forma $(\bar{x}_1 + \alpha_1, \ldots, \bar{x}_n + \alpha_n)$, donde $(\alpha_1, \ldots, \alpha_n)$ es solución de (3.2), su sistema homogéneo asociado.

*Demostración:*
Si $(x'_1, \ldots, x'_n)$ es cualquier otra solución de (3.1) entonces podemos escribir

$$(x'_1, \ldots, x'_n) = (\bar{x}_1, \ldots, \bar{x}_n) + (x'_1 - \bar{x}_1, \ldots, x'_n - \bar{x}_n)$$

Si ahora ponemos $(\alpha_1, \ldots, \alpha_n) = (x'_1 - \bar{x}_1, \ldots, x'_n - \bar{x}_n)$, es inmediato comprobar que esta $n$-upla es solución de (3.2). De manera que hemos podido escribir cualquier solución del sistema no homogéneo como suma de una solución dada, la que se conoce como *solución particular*, más una solución del sistema homogéneo asociado.

■

Como consecuencia de este resultado podemos afirmar que si el sistema (3.1) es compatible, entonces tiene tantas soluciones como su correspondiente sistema homogéneo (3.2).

**Ejemplo 3.2**

Consideremos el sistema (III) del ejemplo 3.1. Su sistema homogéneo asociado es

$$\left.\begin{array}{rcl} x_1 - x_2 + 2x_3 - x_4 & = & 0 \\ -x_1 + 2x_2 - x_3 - x_4 & = & 0 \\ 3x_1 - 4x_2 + 5x_3 - x_4 & = & 0 \end{array}\right\}$$

Si resolvemos el sistema homogéneo mediante el método de Gauss obtendremos como solución general: $(-3\alpha + 3\beta, -\alpha + 2\beta, \alpha, \beta)$.

Imaginemos ahora que conocemos una solución particular del sistema no homogéneo, por ejemplo $(0, -1, 0, 0)$. Cómo llegamos a conocer esta solución particular no es relevante aquí; supongamos que, simplemente, la conocemos. Entonces, según afirma la proposición 3.2 la solución del sistema no homogéneo es

$$(0, -1, 0, 0) + (-3\alpha + 3\beta, -\alpha + 2\beta, \alpha, \beta)$$

que obviamente coincide con la obtenida en el ejemplo 3.1.

### Regla de Cramer

La regla de Cramer[1] nos da la solución directa de un sistema cuadrado compatible determinado mediante el cálculo de ciertos determinantes. Tiene su utilidad pues nos proporciona una expresión *cerrada* de la solución del sistema, aunque en la práctica no es muy útil cuando el sistema es grande pues obliga al cálculo de determinantes de tamaño excesivo.

**Teorema 3.1 (Regla de Cramer)**

Si $A \in \mathcal{M}_n(\mathbb{K})$ es regular y $\mathbf{x}$, $\mathbf{b} \in \mathcal{M}_{n\times 1}(\mathbb{K})$, el sistema $A\mathbf{x} = \mathbf{b}$ es compatible determinado y

$$x_j = \frac{|A_j|}{|A|}, \quad 1 \le j \le n$$

donde $A_j$ es la matriz

$$A_j = \begin{pmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{pmatrix}$$

es decir, la matriz obtenida de $A$ al sustituir la columna $j$ por el segundo miembro del sistema.

*Demostración:*
Como se vio en la nota 2.2, como $A$ es regular, la solución de $A\mathbf{x} = \mathbf{b}$ es $\mathbf{x} = A^{-1}\mathbf{b}$. Usando el Teorema 2.5, $\mathbf{x} = \frac{1}{|A|}\operatorname{adj}(A)^T\mathbf{b}$. Siguiendo con la notación empleada en la demostración del Teorema 2.5, el producto matricial anterior puede escribirse por

$$\mathbf{x} = \frac{1}{|A|}\begin{pmatrix} b_1c_{11} + \cdots + b_nc_{1n} \\ \vdots \\ b_1c_{n1} + \cdots + b_nc_{nn} \end{pmatrix}$$

Es decir,

$$x_j = \frac{1}{|A|}(b_1c_{j1} + \cdots + b_nc_{jn}),$$

expresión que coincide con $\frac{|A_j|}{|A|}$ desarrollado por la $j$-ésima columna. ■

[1]Debe su nombre al matemático suizo Gabriel Cramer que la publicó y popularizó en 1750, aunque el resultado ya había sido publicado dos años antes por el escocés Colin Maclaurin.

**Ejemplo 3.3**

Resolver el siguiente sistema:

$$\left.\begin{array}{rcl} 2x_1 + 3x_2 & = & 8 \\ x_1 + x_3 & = & 1 \\ 2x_2 - x_4 & = & 1 \\ x_1 + x_4 & = & 4 \end{array}\right\}$$

En este caso,

$$A = \begin{pmatrix} 2 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & -1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 8 \\ 1 \\ 1 \\ 4 \end{pmatrix}$$

Desarrollando el determinante de $A$ por la tercera columna obtenemos

$$|A| = -\begin{vmatrix} 2 & 3 & 0 \\ 0 & 2 & -1 \\ 1 & 0 & 1 \end{vmatrix} = -1$$

Puesto que $|A| \neq 0$ el sistema tiene solución única. Calculando los determinantes de las correspondientes matrices $A_j$, $1 \leq j \leq 4$ se tiene,

$$|A_1| = \begin{vmatrix} 8 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & -1 \\ 4 & 0 & 0 & 1 \end{vmatrix} = -\begin{vmatrix} 8 & 3 & 0 \\ 1 & 2 & -1 \\ 4 & 0 & 1 \end{vmatrix} = -1$$

$$|A_2| = \begin{vmatrix} 2 & 8 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 4 & 0 & 1 \end{vmatrix} = -\begin{vmatrix} 2 & 8 & 0 \\ 0 & 1 & -1 \\ 1 & 4 & 1 \end{vmatrix} = -2$$

$$|A_3| = \begin{vmatrix} 2 & 3 & 8 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 4 & 1 \end{vmatrix} \overset{C_3 - C_1}{=} \begin{vmatrix} 2 & 3 & 6 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 3 & 1 \end{vmatrix} = -\begin{vmatrix} 3 & 6 & 0 \\ 2 & 1 & -1 \\ 0 & 3 & 1 \end{vmatrix} = 0$$

$$|A_4| = \begin{vmatrix} 2 & 3 & 0 & 8 \\ 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 0 & 4 \end{vmatrix} = -\begin{vmatrix} 2 & 3 & 8 \\ 0 & 2 & 1 \\ 1 & 0 & 4 \end{vmatrix} = -3$$

Por tanto, $x_1 = 1$, $x_2 = 2$, $x_3 = 0$ y $x_4 = 3$.

## 3 2 TEOREMA DE ROUCHÉ-FROBENIUS

**Teorema 3.2 (Teorema de Rouché-Frobenius[2])**

Consideremos el sistema lineal $A\mathbf{x} = \mathbf{b}$, con $A \in \mathcal{M}_{m\times n}(\mathbb{K})$, $\mathbf{x} \in \mathcal{M}_{n\times 1}(\mathbb{K})$ y $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{K})$. Denotemos por $\bar{A}$ a la matriz *ampliada* del sistema, esto es la matriz formada por $A$ junto con el segundo miembro $\mathbf{b}$. Entonces, el sistema es compatible si y sólo si $\text{rango}(A) = \text{rango}(\bar{A})$.

Más aun, el sistema será determinado si $\text{rango}(A) = n$ (nº de incógnitas) e indeterminado si $\text{rango}(A) < n$.

Nótese además, que en número de grados de libertad del sistema vendrá dado por $n - \text{rango}(A)$.

*Demostración:*

Por el momento sólo probaremos que la condición es necesaria. La prueba de la suficiencia la posponemos hasta el tema siguiente (cf. nota 4.1).

Si el sistema es compatible, entonces deben existir escalares $x_1, \dots, x_n \in \mathbb{K}$ tales que $A\mathbf{x} = \mathbf{b}$, donde, como es evidente, $(\mathbf{x})_i = x_i$. Consideremos ahora la matriz ampliada $\bar{A} \in \mathcal{M}_{m\times(n+1)}$ (la última columna corresponde a la matriz $\mathbf{b}$), y sobre ella efectuamos la operación entre columnas siguiente:

$$C_{n+1} - x_1 C_1 - \cdots - x_n C_n$$

[2] El nombre se debe al matemático francés Eugène Rouché, que lo enunció en 1875 y a Frobenius. Poco después de su publicación, el francés Georges Fontené se atribuyó la primera demostración del mismo, y Frobenius, en 1905, acreditó la demostración a Rouché y Fontené. Sin embargo, el nombre por el que se conoce el resultado (especialmente en los países de habla hispana) se debe al matemático hispano-argentino Julio Rey Pastor, que lo refirió de tal forma. En otros países es conocido como Teorema de Rouché-Fontené o Teorema de Kronecker-Capelli.

Dicha operación da como resultado la matriz

$$M = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & 0 \\ a_{21} & \cdots & a_{2n} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 \end{array}\right)$$

y por la aplicación sucesiva de (IV) del Teorema 2.4 se tiene que $\text{rango}(\bar{A}) = \text{rango}(M)$, que trivialmente es igual al $\text{rango}(A)$.

■

**Ejemplo 3.4**

Discutir la compatibilidad del siguiente sistema mediante el Teorema de Rouché-Frobenius.

$$\left.\begin{array}{rcl} x_1 + 2x_2 & = & 1 \\ 2x_1 + x_2 + 3x_3 & = & 2 \\ 3x_1 + x_2 + 5x_3 & = & 3 \end{array}\right\}$$

La matriz ampliada se escribe,

$$\bar{A} = \left(\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 2 & 1 & 3 & 2 \\ 3 & 1 & 5 & 3 \end{array}\right)$$

Es fácil comprobar que $\text{rango}(A) = 2$, pues

$$\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} \neq 0 \quad \text{y} \quad \begin{vmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 3 & 1 & 5 \end{vmatrix} = 0$$

y $\text{rango}(\bar{A}) = 2$, pues

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 3 & 1 & 3 \end{vmatrix} = 0.$$[3]

Por tanto se trata de un sistema compatible indeterminado ($\text{rango}(A) <$ nº de incógnitas).

Es importante prestar atención a la resolución de este tipo de sistemas. Puesto que $\text{rango}(A) = 2$, esto nos indica que solo dos ecuaciones y dos

incógnitas son relevantes en la resolución del sistema, y por tanto el número de grados de libertad es uno (es decir, el número de incógnitas menos el número de ecuaciones relevantes). Por otra parte, ¿cuáles de estas incógnitas y ecuaciones son las relevantes? La respuesta nos la da el menor básico: es decir, las ecuaciones e incógnitas asociadas a las filas y columnas del menor básico; en este caso, las dos primeras ecuaciones y las dos primeras incógnitas.

Para resolver el sistema, consideramos el resto de incógnitas como parámetros, esto es

$$\left.\begin{array}{rcc} x_1 + 2x_2 & = & 1 \\ 2x_1 + x_2 & = & 2 - 3\alpha \end{array}\right\}$$

donde hemos puesto $x_3 = \alpha$. Ahora tenemos un sistema cuadrado cuya matriz de coeficientes tiene determinante no nulo (ya que se trata del menor básico), así pues podemos usar Cramer en su resolución obteniendo:

$$|A| = -3, \quad |A_1| = \begin{vmatrix} 1 & 2 \\ 2-3\alpha & 1 \end{vmatrix} = 6\alpha - 3, \quad |A_2| = \begin{vmatrix} 1 & 1 \\ 2 & 2-3\alpha \end{vmatrix} = -3\alpha$$

Luego las soluciones del sistema vienen dadas por $(1 - 2\alpha, \alpha, \alpha)$.

**Ejemplo 3.5**

Discutir el sistema

$$\left.\begin{array}{rcc} ax + 2z & = & 2 \\ 5x + 2y & = & 1 \\ x - 2y + bz & = & 3 \end{array}\right\}$$

en función de los valores de $a$ y $b$ y resolverlo cuando sea compatible indeterminado.

La matriz de coeficientes y la matriz ampliada del sistema son

$$A = \begin{pmatrix} a & 0 & 2 \\ 5 & 2 & 0 \\ 1 & -2 & b \end{pmatrix} \quad \bar{A} = \left(\begin{array}{ccc|c} a & 0 & 2 & 2 \\ 5 & 2 & 0 & 1 \\ 1 & -2 & b & 3 \end{array}\right)$$

Para estudiar el rango de $A$ vemos que las dos primeras filas y las dos últimas columnas conforman un menor de orden dos distinto de cero (con independencia

de los valores de $a$ y $b$), de modo que al menos el rango de $A$ es dos. Estudiando

$$|A| = \begin{vmatrix} a & 0 & 2 \\ 5 & 2 & 0 \\ 1 & -2 & b \end{vmatrix} = 2ab - 24$$

observamos que los valores de $a$ y $b$ que hacen que $|A| = 0$ son los que verifican $ab = 12$. En conclusión, si $ab = 12$, rango$(A) = 2$ y en caso contrario rango$(A) = 3$.

Veamos ahora qué pasa con $\bar{A}$: obviamente, si $ab \neq 12$, como $\bar{A}$ no puede tener rango cuatro, rango$(\bar{A}) = 3$ y el sistema será compatible determinado. Nos resta por ver qué pasa si $ab = 12$. Orlando el menor de orden dos de $A$ con la última columna de $\bar{A}$,

$$\begin{vmatrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ -2 & b & 3 \end{vmatrix} = 4b - 16$$

De modo que, si $b = 4$ (y como $ab = 12$ eso significa que $a = 3$), entonces rango$(\bar{A}) = 2$ y el sistema es compatible indeterminado. Si $b \neq 4$, (y $ab = 12$) el sistema es incompatible. Resumiendo

$$\begin{cases} ab \neq 12 & \text{sistema compatible determinado} \\ b = 4,\ a = 3 & \text{sistema compatible indeterminado} \\ \text{en otro caso} & \text{sistema incompatible} \end{cases}$$

Resolvamos ahora para $a = 3$ y $b = 4$. Consideramos el menor de orden dos escogido previamente y usamos la incógnita correspondiente a la columna que no aporta rango como parámetro (en este caso la primera, o sea, la variable $x$); pasando los parámetros al segundo miembro,

$$\left.\begin{aligned} 2z &= 2 - 3\alpha \\ 2y &= 1 - 5\alpha \end{aligned}\right\}$$

La solución es $\left(\alpha, \frac{1-5\alpha}{2}, \frac{2-3\alpha}{2}\right)$.

3 3

## ÁLGEBRA LINEAL NUMÉRICA

Es muy frecuente que un gran número de problemas matemáticos provenientes de diversos campos precisen, en algún momento de su resolución, tener que tratar con sistemas lineales con un gran número de ecuaciones e incógnitas. El tamaño de tales sistemas los vuelve imposibles de resolver a mano, por lo que es necesario acudir a algoritmos numéricos programados en un ordenador para obtener su solución. El Álgebra Lineal Numérica se encarga de estudiar y analizar métodos para resolver estos sistemas computacionalmente.

Básicamente se distinguen dos tipos de métodos: *directos* e *indirectos*. Los primeros tratan de obtener la solución del sistema tras un número finito de pasos, mientras que los segundos construyen una sucesión de valores que van convergiendo a la solución del sistema. Antes de pasar a describir algoritmos de ambos métodos hemos de prestar atención a la forma en la que los ordenadores trabajan con los números, pues es crucial en determinados aspectos de los métodos numéricos.

### Errores de redondeo

Los ordenadores trabajan con la denominada *aritmética de coma flotante* en la que cada número es representado en la máquina de una forma determinada que permite almacenar números reales extremadamente grandes y pequeños de forma compacta. Para nosotros, lo más importante de este hecho es que las máquinas tienen que trabajar con un número limitado de cifras decimales. Por ejemplo, una computadora no puede[4] tratar con el número $\frac{8}{3} = 2.\widehat{6}$, pues dependiendo del *número de cifras significativas* que maneje la máquina, el número se almacenará solo con una cantidad finita de cifras, lo cual induce un determinado error. Estos errores se pueden producir, bien por *truncamiento*, es decir, en lugar de considerar $\frac{8}{3}$ consideramos 2.6666, (si nuestra máquina considerara sólo cuatro cifras significativas) truncando la expansión decimal, o bien por *redondeo*, en el que almacenaríamos el dígito de cuatro cifras decimales más próximo al número en cuestión, es decir, en este caso almacenaríamos 2.6667. En ambos casos hablamos de errores de redondeo.

Si bien los errores de redondeo no parecen ser importantes, más aun cuando el número de cifras significativas con el que trabaja una computadora actualmente se sitúa en 16 ó 32, hay que tener cuidado con los mismos pues pueden aparecer diferencias notables. Por ejemplo, si calculamos exactamente los valores

$$\frac{1}{\frac{8}{3} - 2.6666} = 15000, \qquad \frac{1}{\frac{8}{3} - 2.6667} = -30000$$

[4]Esto no es totalmente cierto: hemos visto cómo el módulo SymPy de Python permite trabajar con números como si fueran símbolos, y por tanto es posible realizar ciertos cálculos de forma exacta.

¡la diferencia entre ambos valores es de 45000! (es lo que se denomina el *error absoluto*, que corresponde al valor absoluto de la diferencia entre los valores). En este caso, tal disparidad de valores se debe al hecho de que los denominadores de las fracciones calculadas son muy pequeños. En tales casos, una pequeña diferencia de valores en los denominadores provoca diferencias enormes en los resultados. Uno de los retos del Álgebra Lineal Numérica consiste en el diseño métodos que reduzcan en la medida de lo posible los errores generados por el redondeo, así como tratar de estimar estos errores cuando ocurren.

Medir el error cometido mediante los errores absolutos hace que perdamos perspectiva. Por ejemplo, si estamos midiendo la longitud de un tornillo cuyo valor real es de 5 mm., y el aparato de medición nos proporciona 4 mm. estamos cometiendo un error absoluto de 1 mm. Si la medición es la de un campo de fútbol que tiene 100 m. de largo, y nuestra medición es de 99 m. estaremos cometiendo un error absoluto de 1 m. Es evidente que el error absoluto en el segundo caso es mucho mayor que en el primero, pero ¿cuál de las dos medidas es mejor? En el primer caso nos equivocamos en 1 mm. de cada 5 mm., mientras que en el segundo, el error es de 1 m. de cada 100 m. Es decir, es necesario *relativizar* el error en función del tamaño de las cantidades con las que estamos trabajando. Por eso se trabaja con el *error relativo* que corresponde al cociente entre el error absoluto y la cantidad a aproximar. Así pues, en la medición del tornillo estamos cometiendo un error de $\frac{1}{5} = 0.2$, mientras que al medir el campo de fútbol hemos cometido un error relativo de $\frac{1}{100} = 0.01$ (bastante menor que el anterior). Es habitual usar porcentajes cuando trabajamos con errores relativos, de modo que, teniendo en cuenta que éstos vienen dados en tantos por uno, multiplicamos por 100 para obtener el tanto por cien de error (un 20 % para el primer caso y un 1 % para el segundo).

### 3 3 1 Métodos directos

El método directo más conocido para la resolución de sistemas es el método de Gauss que vimos en la sección 2.2. Con este método, tras un número finito de pasos llegamos a la solución exacta del sistema, siempre y cuando no cometamos errores de redondeo. Tales errores, tal y como acabamos de ver, pueden tener consecuencias importantes, de ahí que se hayan desarrollado estrategias encaminadas a minimizar estos efectos no deseados.

Para ver de qué manera afectan los errores de redondeo al método de Gauss, veamos un ejemplo en el que los cálculos los haremos de forma similar a como los realiza un computador. Para ello será necesario explicar de forma muy resumida cómo funciona la aritmética de coma flotante. Básicamente se trata de representar los números reales mediante una expresión del tipo

$$x = m \cdot 10^e$$

donde $m$ es la denominada *mantisa* y $e$ es el *exponente*.[5] La mantisa es de la

[5] En realidad los ordenadores trabajan en base 2, pero por simplicidad en la exposición,

forma $m = 0.d_1d_2\cdots d_k$ con $d_1 \neq 0$, mientras que $k$ es el número de cifras significativas. Así, si por ejemplo, el computador que estamos usando emplea 4 cifras significativas, el número 22.385 se representa por $0.2238\cdot 10^2$, mientras que 0.00023 se representa por $0.23\cdot 10^{-3}$. La operación suma (o resta) de dos números en coma flotante es realizada por los computadores de una forma peculiar: si queremos sumar $x = m_1 \cdot 10^{e_1}$ con $y = m_2 \cdot 10^{e_2}$, con $e_1 > e_2$, haremos lo siguiente

$$x + y = m_1 \cdot 10^{e_1} + m_2 \cdot 10^{e_2} = (m_1 + m_2 \cdot 10^{e_2 - e_1}) \cdot 10^{e_1}$$

Por ejemplo,

$$\begin{aligned} 22.385 + 0.00023 &= 0.2238 \cdot 10^2 + 0.23 \cdot 10^{-3} \\ &= (0.2238 + 0.23 \cdot 10^{-5}) \cdot 10^2 = 0.2238023 \cdot 10^2 \end{aligned}$$

que con 4 cifras significativas es $0.2238 \cdot 10^2$.

Veamos cómo afectan estos cálculos a la resolución de un sistema:

**Ejemplo 3.6**

Consideremos el siguiente sistema:

$$\left.\begin{array}{rcl} 0.0001x_1 + x_2 &=& 2 \\ x_1 + x_2 &=& 3 \end{array}\right\}$$

cuya solución exacta es $x_1 = \frac{10000}{9999} \approx 1.0001$ y $x_2 = \frac{19997}{9999} \approx 1.9999$. Supongamos que nuestro computador trabaja con aritmética de coma flotante con 4 cifras significativas. Las operaciones del método de Gauss son:

$$\left(\begin{array}{cc|c} 0.1 \cdot 10^{-3} & 0.1 \cdot 10 & 0.2 \cdot 10 \\ 0.1 \cdot 10 & 0.1 \cdot 10 & 0.3 \cdot 10 \end{array}\right) \xrightarrow{F_2 - 10^4 F_1} \left(\begin{array}{cc|c} 0.1 \cdot 10^{-3} & 0.1 \cdot 10 & 0.2 \cdot 10 \\ 0 & -0.9999 \cdot 10^5 & -0.2 \cdot 10^5 \end{array}\right)$$

Resolviendo queda $x_1 = 0$ y $x_2 = 2$, de modo que la solución numérica para $x_1$ tiene un error relativo aproximado del 100 %.

¿A qué se debe esta enorme diferencia? Observemos que en este ejemplo, la única operación que lleva a cabo el método de Gauss es $F_2 - \frac{1}{0.0001}F_1$. Como podemos apreciar, estamos dividiendo por un número cercano a cero, lo cual, en virtud del comentario hecho antes, distorsiona los resultados.

nosotros usamos la base 10.

¿Qué ocurre si en lugar de resolver el sistema tal y como lo tenemos, lo resolvemos intercambiando previamente las ecuaciones? En este caso

$$\left(\begin{array}{cc|c} 0.1\cdot 10 & 0.1\cdot 10 & 0.3\cdot 10 \\ 0.1\cdot 10^{-3} & 0.1\cdot 10 & 0.2\cdot 10 \end{array}\right)$$

$$\xrightarrow{F_2-10^{-4}F_1} \left(\begin{array}{cc|c} 0.1\cdot 10 & 0.1\cdot 10 & 0.3\cdot 10 \\ 0 & 0.0999\cdot 10 & 0.1997\cdot 10 \end{array}\right)$$

Ahora las soluciones son $x_1 = 1.0009$ y $x_2 = 1.9990$. Vemos ahora que el mayor error relativo cometido no llega al $0.1\,\%$.

Si revisamos con atención las operaciones que realiza el método de Gauss sobre un sistema observamos que, en cada iteración, con objeto de triangular el sistema, es preciso multiplicar por el cociente entre el elemento de la columna que queremos anular y el elemento de la diagonal que usamos como referencia. Este elemento es el denominado *pivote*. En el primer caso del ejemplo 3.6, el pivote era 0.0001, que al ser próximo a cero hace que las divisiones por éste conlleven errores mayores de lo que cabría esperar. Por el contrario, en el segundo caso del mismo ejemplo, al intercambiar las ecuaciones entre sí, el pivote es ahora 1, y no se producen distorsiones al dividir por él.

Puesto que el intercambio de filas en la matriz de un sistema no altera las soluciones del mismo, el conocido como *método de Gauss con estrategia de pivoteo parcial*, consiste sencillamente en intercambiar las ecuaciones de orden con objeto de que el pivote resulte ser el de mayor valor absoluto de entre todos los posibles, que es lo que se ha hecho en el ejemplo 3.6. Veámoslo en un nuevo ejemplo.

**Ejemplo 3.7**

Resolvamos el siguiente sistema usando el método de Gauss con y sin estrategia de pivoteo parcial usando aritmética de coma flotante con 4 cifras significativas y comparemos resultados

$$\left.\begin{array}{rcr} -2x+4y-16z & = & -38 \\ 14x+2y+4z & = & -10 \\ 16x+40y-4z & = & 55 \end{array}\right\}$$

Nótese que la solución exacta es $x = 2$, $y = 2.5$, $z = 3.25$.

Sin estrategia de pivoteo,

$$\left(\begin{array}{rrr|r} -2 & 4 & -16 & -38 \\ 14 & 2 & 4 & -10 \\ 16 & 40 & -4 & 55 \end{array}\right)$$

$$\xrightarrow[F_3+8F_1]{F_2+7F_1} \left(\begin{array}{ccc|c} -0.2\cdot 10 & 0.4\cdot 10 & -0.16\cdot 10^2 & -0.38\cdot 10^2 \\ 0 & 0.3\cdot 10^2 & -0.108\cdot 10^3 & -0.276\cdot 10^3 \\ 0 & 0.72\cdot 10^2 & -0.132\cdot 10^3 & -0.249\cdot 10^3 \end{array}\right)$$

$$\xrightarrow{F_3-2.4F_2} \left(\begin{array}{ccc|c} -0.2\cdot 10 & 0.4\cdot 10 & -0.16\cdot 10^2 & -0.38\cdot 10^2 \\ 0 & 0.3\cdot 10^2 & -0.108\cdot 10^3 & -0.276\cdot 10^3 \\ 0 & 0 & 0.1272\cdot 10^3 & 0.413\cdot 10^3 \end{array}\right)$$

Resolviendo obtenemos: $x = 2.033$, $y = 2.4666$ y $z = 3.2468$.

Procedamos ahora usando el método de Gauss con estrategia de pivoteo parcial. En lugar de escoger como pivote el $-2$, usaremos el de mayor valor absoluto de la primera columna, esto es 16, por lo que procedemos en primer lugar intercambiando las filas primera y tercera,

$$\left(\begin{array}{rrr|r} -2 & 4 & -16 & -38 \\ 14 & 2 & 4 & -10 \\ 16 & 40 & -4 & 55 \end{array}\right) \xrightarrow{F_1\leftrightarrow F_3} \left(\begin{array}{rrr|r} 16 & 40 & -4 & 55 \\ 14 & 2 & 4 & -10 \\ -2 & 4 & -16 & -38 \end{array}\right)$$

$$\xrightarrow[F_3+0.125F_1]{F_2-0.875F_1} \left(\begin{array}{ccc|c} 0.16\cdot 10^2 & 0.4\cdot 10^2 & -0.4\cdot 10 & 0.55\cdot 10^2 \\ 0 & -0.33\cdot 10^2 & 0.75\cdot 10 & -0.5812\cdot 10^2 \\ 0 & 0.9\cdot 10 & -0.165\cdot 10^2 & -0.3112\cdot 10^2 \end{array}\right)$$

El siguiente paso será elegir el pivote para la segunda columna. Debemos mirar cuál, de entre los elementos de la segunda columna que quedan por debajo de la diagonal (no es posible escoger como pivote el de la primera fila, pues perderíamos el cero ya obtenido), tiene mayor valor absoluto. Observamos que ahora no es necesario intercambiar filas pues el pivote de mayor valor absoluto ya se encuentra en la segunda fila. Así:

$$\xrightarrow{F_3+0.2727F_2} \left(\begin{array}{ccc|c} 0.16\cdot 10^2 & 0.4\cdot 10^2 & -0.4\cdot 10 & 0.55\cdot 10^2 \\ 0 & -0.33\cdot 10^2 & 0.75\cdot 10 & -0.5812\cdot 10^2 \\ 0 & 0 & -0.1445\cdot 10^2 & -0.4696\cdot 10^2 \end{array}\right)$$

y resolviendo obtenemos $x = 1.9981$, $y = 2.4993$ y $z = 3.2498$.

Vemos que la aproximación en el segundo caso es un poco mejor que en el primero. ¿Puede el lector calcular los errores relativos de las dos aproximaciones?

En definitiva, con la técnica del pivoteo parcial se pretende que los errores de redondeo cometidos de forma natural por las computadoras queden minimizados. Existe una versión similar al pivoteo parcial, denominada *método de Gauss con estrategia de pivoteo total* que, en lugar de elegir como pivote el elemento de mayor valor absoluto de la columna, selecciona el de mayor valor absoluto considerando todas las filas y columnas a partir del elemento diagonal que corresponda. Si bien computacionalmente el pivoteo total proporciona mejores resultados, su implementación precisa de un renombrado de variables (cuando nos vemos obligados a intercambiar columnas entre sí) que lo hacen menos popular que la técnica del pivoteo parcial.

### Factorización LU

Es frecuente en las aplicaciones que tengamos que resolver una cierta cantidad de sistemas del tipo $A\mathbf{x} = \mathbf{b}$ en los que la matriz $A$ permanece fija, pero el segundo miembro cambia (véase la sección 3.5). La *factorización $LU$* consiste en escribir la matriz $A$ como el producto de dos matrices, una triangular inferior $L$ y una triangular superior $U$ de manera que $A = LU$.[6]

Si conocemos una factorización de la matriz $A$ de este tipo y queremos resolver el sistema $A\mathbf{x} = \mathbf{b}$, entonces podemos proceder del siguiente modo:

$$A\mathbf{x} = \mathbf{b} \iff LU\mathbf{x} = \mathbf{b} \iff \begin{cases} L\mathbf{y} = \mathbf{b} \\ U\mathbf{x} = \mathbf{y} \end{cases}$$

Es decir, en primer lugar resolvemos un sistema con matriz $L$ y segundo miembro $\mathbf{b}$, denotando por $\mathbf{y}$ su solución, y a continuación, resolvemos el sistema con matriz $U$ y segundo miembro $\mathbf{y}$, cuya solución será $\mathbf{x}$, la solución del sistema original.

¿Qué ventaja supone resolver los sistemas $L\mathbf{y} = \mathbf{b}$ y luego $U\mathbf{x} = \mathbf{y}$ en lugar del sistema original $A\mathbf{x} = \mathbf{b}$? Si observamos las matrices de estos sistemas, $L$ es una matriz triangular inferior, de manera que la resolución de un sistema con esta matriz se reduce a un sencillo y rápido proceso de subida. Por su parte, $U$ es triangular superior de manera que el sistema $U\mathbf{x} = \mathbf{y}$ se resuelve fácilmente

[6] El nombre proviene de las siglas en inglés L por *lower* (inferior) y U por *upper* (superior). Aunque el método básico ya era conocido por Gauss, fue el norteamericano Myrick Hascall Doolittle a finales del s. XIX quien lo formuló de manera algorítmica.

mediante una *bajada*.[7] En definitiva, una vez conocida la factorización LU de una matriz $A$, cualquier sistema se va a poder resolver realizando primero un proceso de subida y luego uno de bajada, lo cual es enormemente ventajoso desde el punto de vista computacional.

El cálculo de la factorización LU se lleva a cabo de forma sencilla a través del método de Gauss y de las matrices elementales introducidas en la sección 2.3. Recordemos que las matrices elementales permiten expresar las transformaciones que se llevan a cabo en el método de Gauss como producto de matrices.

Veámoslo con un ejemplo.

**Ejemplo 3.8**

Si aplicamos el método de Gauss a la matriz

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

olvidándonos del segundo miembro, las operaciones elementales sobre esta matriz se escriben,

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{F_2-3F_1} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & -1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{F_3+\frac{1}{5}F_2} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -5 & -1 \\ 0 & 0 & \frac{4}{5} \end{pmatrix} = U$$

Nótese que la matriz final es triangular superior y la denotamos por $U$. Cada una de las operaciones realizadas tiene su correspondiente matriz elemental:

$$F_2 - 3F_1 \longrightarrow E_{12}(-3) = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$F_3 + \tfrac{1}{5}F_2 \longrightarrow E_{23}(\tfrac{1}{5}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{5} & 1 \end{pmatrix}$$

Es decir, en términos de matrices elementales, el método de Gauss ha realizado la multiplicación

$$E_{23}(\tfrac{1}{5})E_{12}(-3)A = U$$

[7] Esto es lo análogo de una subida, sólo que en lugar de empezar por la última ecuación, comenzamos por la primera.

Nótese además que cada una de las matrices elementales empleadas es triangular inferior.[8] Además, el producto de matrices triangulares inferiores es una matriz triangular inferior, y la inversa de una triangular inferior, también es triangular inferior. Así pues, la matriz $E_{23}(\frac{1}{5})E_{12}(-3)$ y su inversa son matrices triangulares inferiores. Por tanto, podemos escribir

$$A = \left(E_{23}(\tfrac{1}{5})E_{12}(-3)\right)^{-1} U = (E_{12}(-3))^{-1} \left(E_{23}(\tfrac{1}{5})\right)^{-1} U$$

Denotando por $L = (E_{12}(-3))^{-1} \left(E_{23}(\frac{1}{5})\right)^{-1}$, que es una matriz triangular inferior, obtenemos la factorización LU de $A$. Es decir, hemos escrito $A$ como un producto de una matriz triangular inferior y una matriz triangular superior.

En realidad, para calcular la matriz $L$ no es necesario calcular las inversas de las matrices elementales y luego los productos entre las mismas, pues estas operaciones se pueden hacer de forma más simple. Por una parte, las inversas de las matrices elementales son inmediatas; más concretamente se tiene que

$$(E_{ij}(\lambda))^{-1} = E_{ij}(-\lambda), \qquad (E_i(\lambda))^{-1} = E_i(\tfrac{1}{\lambda})$$

Es decir, las inversas de matrices elementales son también matrices elementales. Por tanto, efectuar el producto de tales matrices equivale a realizar la correspondiente operación que representan sobre la matriz identidad (atención al orden de la multiplicación). En el caso que nos ocupa:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{E_{23}(-\frac{1}{5})} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{5} & 1 \end{pmatrix} \xrightarrow{E_{12}(3)} \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & -\frac{1}{5} & 1 \end{pmatrix} = L$$

obteniéndose de este modo la matriz $L$ buscada.

Más aun, calcular la matriz $L$ no requiere tener en cuenta todo lo anterior, sino simplemente prestar atención a las operaciones llevadas a cabo en el método de Gauss. Lo vemos en el siguiente ejemplo.

[8] Este hecho se cumple para todas las matrices elementales de tipo (ii) y (iii), siempre que $i < k$, pero no es cierto para las de tipo (i) (véase la Definición 2.11).

**Ejemplo 3.9**

Resolver mediante una factorización LU el sistema siguiente:

$$\left.\begin{array}{rcl} x_1 - x_2 + 2x_3 & = & 2 \\ 2x_1 + x_2 + x_3 & = & 3 \\ 3x_1 - x_2 + x_3 & = & 3 \end{array}\right\}$$

Llevaremos a cabo en primer lugar la factorización LU de la matriz de coeficientes:

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 3 & -1 & 1 \end{pmatrix}$$

La operaciones a realizar son:

$$\begin{pmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 3 & -1 & 1 \end{pmatrix} \xrightarrow[]{\substack{F_2-2F_1 \\ F_3-3F_1}} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & -3 \\ 0 & 2 & -5 \end{pmatrix} \xrightarrow{\frac{1}{3}F_2} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & 2 & -5 \end{pmatrix}$$

$$\xrightarrow{F_3-2F_2} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & -3 \end{pmatrix} = U$$

Para construir la matriz $L$ simplemente "coleccionamos" las operaciones "inversas" de las realizadas en la matriz identidad:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[\substack{\mathbf{F_2+2F_1} \\ \mathbf{F_3+3F_1}}]{\substack{F_2-2F_1 \\ F_3-3F_1}} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \xrightarrow[\mathbf{3F_2}]{\frac{1}{3}F_2} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[\mathbf{F_3+2F_2}]{F_3-2F_2} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 3 & 2 & 1 \end{pmatrix} = L$$

Obsérvese que hemos escrito encima la operación que marcaba el método de Gauss, y debajo y en negrita, la operación inversa que hemos realizado.

Podemos comprobar fácilmente que $LU = A$. Para resolver ahora el sistema inicial resolvemos primero el sistema $L\mathbf{y} = \mathbf{b}$. Es decir,

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$$

cuya solución es $y_1 = 2$, $y_2 = -\frac{1}{3}$, $y_3 = -\frac{7}{3}$; luego resolvemos $U\mathbf{x} = \mathbf{y}$, es decir,

$$\begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -\frac{1}{3} \\ -\frac{7}{3} \end{pmatrix}$$

con solución $x_1 = \frac{8}{9}$, $x_2 = \frac{4}{9}$ y $x_3 = \frac{7}{9}$.

Es importante resaltar el hecho de que la factorización LU de una matriz no es única, puesto que el método de Gauss tampoco lo es. Por otra parte, hay que mencionar que en determinados sistemas el proceso llevado a cabo no es posible. Por ejemplo, si el elemento $a_{11} = 0$, para aplicar el método de Gauss es preciso intercambiar filas de la matriz, lo cual lleva consigo la introducción de una matriz elemental que ya no es triangular, lo que nos conduciría a una matriz $L$ no triangular inferior. En realidad esto no supone un problema grave, pues el remedio consiste en intercambiar desde el principio el orden de las filas de la matriz (es decir, intercambiar el orden de las ecuaciones) para no tener que introducir matrices no triangulares en el proceso. En tales casos, las matrices $L$ y $U$ obtenidas no son una factorización de la matriz original, sino de una matriz en la que se han intercambiado algunas de sus filas. Veámoslo en el siguiente ejemplo.

**Ejemplo 3.10**

Encontrar la factorización LU de la matriz

$$A = \begin{pmatrix} 0 & 2 & 5 \\ 2 & 4 & 3 \\ 2 & 1 & 1 \end{pmatrix}$$

Observemos cómo el primer pivote no nos sirve, lo cual nos obliga a cambiar el orden de las filas de la matriz. Así, aplicamos el método de Gauss a la matriz

$$A' = \begin{pmatrix} 2 & 4 & 3 \\ 0 & 2 & 5 \\ 2 & 1 & 1 \end{pmatrix} \xrightarrow{F_3 - F_1} \begin{pmatrix} 2 & 4 & 3 \\ 0 & 2 & 5 \\ 0 & -3 & -2 \end{pmatrix} \xrightarrow{F_3 + \frac{3}{2}F_2} \begin{pmatrix} 2 & 4 & 3 \\ 0 & 2 & 5 \\ 0 & 0 & \frac{11}{2} \end{pmatrix} = U$$

mientras que

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[\mathbf{F_3+F_1}]{F_3-F_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow[\mathbf{F_3-\frac{3}{2}F_2}]{F_3+\frac{3}{2}F_2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -\frac{3}{2} & 1 \end{pmatrix} = L$$

Se tiene por tanto que $LU = A'$ o también $LU = PA$ donde $P$ es la matriz correspondiente al intercambio entre las filas uno y dos, es decir

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Este tipo de matrices es conocido como matriz de *permutación*.

### 3 3 2 Sistemas mal condicionados

Hemos visto al comienzo de esta sección que cuando tratamos de resolver un sistema lineal numéricamente es preciso tener en cuenta la aparición de errores de redondeo provocados por la aritmética de coma flotante con la que trabajan las computadoras. Pero no son éstos los únicos errores con los que debemos contar. Es frecuente que en el planteamiento de un sistema lineal, la recogida de datos que conduce al mismo esté afectada por errores de precisión, es decir, que en lugar de estar resolviendo un sistema $A\mathbf{x} = \mathbf{b}$, tenemos un sistema ligeramente distinto $\tilde{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$, denominado generalmente *sistema perturbado*. La cuestión en este caso es ver si el hecho de que la diferencia entre el sistema original y el perturbado sea pequeña conduce a que la diferencia entre sus soluciones también lo sea. Veamos un ejemplo.

**Ejemplo 3.11**

Consideremos el sistema

$$\left.\begin{array}{rcl} 2x + 3y & = & 5 \\ (2+10^{-7})x + 3y & = & 5 \end{array}\right\}$$

cuya solución es $x = 0$, $y = \frac{5}{3}$. Modificamos ligeramente el sistema anterior para

tener:

$$\left.\begin{array}{rcl} 2x+3y & = & 5 \\ (2+10^{-7})x+3y & = & 5+10^{-7} \end{array}\right\}$$

La solución exacta es ahora $x = 1$, $y = 1$. Obsérvese que la modificación realizada al sistema es mínima (estamos perturbando sólo el segundo miembro muy ligeramente), sin embargo la solución del sistema original es muy distinta a la del sistema perturbado.

El anterior es un típico ejemplo de lo que se conoce como un sistema *mal condicionado*,[9] en oposición a lo que se denomina un sistema bien condicionado, en el que pequeñas perturbaciones en los datos del sistema conducen a pequeñas variaciones en su solución. En lo que sigue veremos que el condicionamiento de un sistema depende esencialmente de un parámetro denominado *número de condición*.

Para hablar del número de condición es preciso hablar antes de *normas*, concepto que introduciremos detenidamente en el tema 8. Por el momento nos bastará tener una idea intuitiva de lo que es una norma vectorial. Esencialmente una norma es una "medida" del tamaño de un vector;[10] más concretamente, se define la norma de un vector $\mathbf{x}$, que se notará como $\|\mathbf{x}\|$, como un número real que verifica las siguientes propiedades:

(I) $\|\mathbf{x}\| \geq 0$, y $\|\mathbf{x}\| = 0$ si y solo si $\mathbf{x} = \mathbf{0}$.

(II) $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|$, $\forall\lambda \in \mathbb{K}$.

(III) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Para vectores en $\mathbb{K}^n$, el ejemplo más conocido es la norma *euclídea* o *norma 2*, definida por

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}, \quad \text{si } \mathbf{x} = (x_1, \dots, x_n)$$

Otras normas habituales son la *norma 1*:

$$\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|$$

y la *norma infinito* o *norma del supremo*:

$$\|\mathbf{x}\|_\infty = \text{máx}\{|x_1|, \dots, |x_n|\}$$

---

[9] El concepto de condicionamiento fue introducido por el matemático inglés Alan Turing en 1948 en un trabajo titulado *Rounding-off errors in matrix processes*, convirtiéndose en el fundador del Álgebra Lineal Numérica moderna.

[10] Hablaremos de vectores a partir del tema siguiente aunque seguramente el lector ya esté familiarizado con ellos.

**Definición 3.3**

Sean $\|\cdot\|_\spadesuit$ y $\|\cdot\|_\clubsuit$ dos normas vectoriales en $\mathbb{K}^m$ y $\mathbb{K}^n$, respectivamente. Se denomina *norma matricial* en $\mathcal{M}_{m\times n}(\mathbb{K})$ *inducida* por dichas normas vectoriales a

$$\|A\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_\spadesuit}{\|\mathbf{x}\|_\clubsuit}$$

En general, es habitual que las normas vectoriales consideradas sean la misma (aunque en diferentes espacios) por lo que evitaremos los subíndices en las normas para no recargar la notación.

Es fácil, a partir de la definición, probar que una norma matricial satisface las propiedades (I)–(III) que definen una norma. Por otra parte, es inmediato comprobar que

$$\|A\mathbf{x}\| \leq \|A\|\,\|\mathbf{x}\|, \quad \forall \mathbf{x} \in \mathbb{K}^{\mathbf{n}} \tag{3.3}$$

También se puede probar que si $\|\cdot\|$ es una norma en $\mathcal{M}_n(\mathbb{K})$ inducida por normas vectoriales, entonces

$$\|AB\| \leq \|A\|\,\|B\|, \quad \forall A, B \in \mathcal{M}_n(\mathbb{K})$$

**Definición 3.4**

Sea $A \in \mathcal{M}_n(\mathbb{K})$ regular y $\|\cdot\|$ una norma matricial inducida por una norma vectorial. Se define el *número de condición* de $A$ como

$$\kappa(A) = \|A\|\,\|A^{-1}\|$$

Si $A$ es singular, se define $\kappa(A) = \infty$

La definición se puede extender para el caso de matrices no cuadradas, pero no la usaremos aquí. El número de condición de una matriz nos da una medida del condicionamiento del sistema: cuanto mayor es, peor es el condicionamiento. Más concretamente, supongamos que tenemos un sistema cuadrado con matriz regular $A\mathbf{x} = \mathbf{b}$ y el correspondiente sistema perturbado $A\tilde{\mathbf{x}} = \mathbf{b} + \Delta\mathbf{b}$, en el que consideramos una variación del segundo miembro "pequeña". La idea es analizar la diferencia entre la solución del sistema original $\mathbf{x}$ y la del sistema perturbado $\tilde{\mathbf{x}}$.

Restando ambos sistemas:

$$\left.\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A\tilde{\mathbf{x}} &= \mathbf{b} + \Delta\mathbf{b} \end{aligned}\right\} \Longrightarrow A(\tilde{\mathbf{x}} - \mathbf{x}) = \Delta\mathbf{b}$$

y puesto que $A$ es invertible, multiplicando por $A^{-1}$ y usando (3.3)

$$\tilde{\mathbf{x}} - \mathbf{x} = A^{-1}\Delta\mathbf{b} \Longrightarrow \|\tilde{\mathbf{x}} - \mathbf{x}\| \leq \|A^{-1}\|\, \|\Delta\mathbf{b}\|$$

Por otra parte (usando nuevamente (3.3)), $\|A\mathbf{x}\| = \|\mathbf{b}\| \Rightarrow \|\mathbf{x}\| \geq \dfrac{\|\mathbf{b}\|}{\|A\|}$. Luego,

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{1}\|\, \|\Delta\mathbf{b}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\|\, \|A\| \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = \kappa(A)\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{3.4}$$

La cantidad $\|\tilde{\mathbf{x}} - \mathbf{x}\|$ es el error absoluto entre la solución del sistema original y la solución del sistema perturbado, mientras que el cociente $\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|}$ es el error relativo. Por su parte, $\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$ es el error relativo de la perturbación que hacemos del sistema. En consecuencia, lo que (3.4) pone de manifiesto es que el error relativo al resolver un sistema está acotado por el número de condición de la matriz multiplicado por el error relativo de la perturbación. Si se supone que la perturbación es pequeña y el número de condición también, el error cometido será pequeño. Si por el contrario, el número de condición es grande no podemos esperar a priori que el error cometido sea pequeño (aunque en algún caso concreto sí que lo pueda ser).

En el ejemplo 3.11, lo que está ocurriendo es que el número de condición de la matriz del mismo es grande. Nos remitimos a la sección 3.4 en la que calculamos el número de condición para esta matriz.

### 3 3 3 Métodos iterativos

Hemos hablado ya de los métodos directos para resolver sistemas, que son aquéllos con los que se pretende llegar a su solución tras un número finito de pasos. En esta sección presentaremos un par de métodos iterativos que construyen una sucesión de valores que deben tender hacia la solución del sistema. Cuando esto ocurre diremos que el método *converge* y en caso contrario que *diverge*. El funcionamiento de los métodos iterativos es siempre el mismo, solo cambia el modo en el que se construye cada iteración. Así, se parte de una aproximación inicial $\mathbf{x}^{(0)}$, donde el superíndice hará referencia al elemento de la sucesión, a partir de la cual debemos poder construir un nuevo elemento $\mathbf{x}^{(1)}$ que se aproxime aun más a la solución del sistema, reiterándose el procedimiento un determinado número de veces. Si el método converge, al cabo de un cierto número de pasos el término de la sucesión obtenido debería ser una buena aproximación a la solución buscada. Este tipo de métodos es bastante eficaz para matrices grandes *vacías*, esto es, con una gran cantidad de ceros.

### Método de Jacobi [11]

Supongamos que tenemos el sistema de $n$ ecuaciones con $n$ incógnitas siguiente

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = & b_n \end{array}\right\}$$

en el que todos los elementos diagonales $a_{ii} \neq 0$. Dada una aproximación $\mathbf{x}^{(k)}$, se construye $\mathbf{x}^{(k+1)}$ del siguiente modo:

$$\begin{array}{rcl} x_1^{(k+1)} & = & \dfrac{1}{a_{11}}\left(b_1 - a_{12}x_2^{(k)} - \cdots - a_{1n}x_n^{(k)}\right) \\ x_2^{(k+1)} & = & \dfrac{1}{a_{22}}\left(b_2 - a_{21}x_1^{(k)} - \cdots - a_{2n}x_n^{(k)}\right) \\ & \vdots & \\ x_n^{(k+1)} & = & \dfrac{1}{a_{nn}}\left(b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \cdots - a_{n,n-1}x_{n-1}^{(k)}\right) \end{array} \tag{3.5}$$

Es decir, básicamente obtenemos la componente i-ésima de la nueva aproximación resolviendo la ecuación i-ésima en esa misma variable. Dicho de otro modo, en cada iteración resolvemos $n$ ecuaciones, cada una de ellas con una única incógnita.

**Ejemplo 3.12**

Aplicar el método de Jacobi al sistema

$$\left.\begin{array}{rcr} 10x_1 - x_2 + 2x_3 & = & 6 \\ -x_1 + 11x_2 - x_3 + 3x_4 & = & 25 \\ 2x_1 - x_2 + 10x_3 - x_4 & = & -11 \\ 3x_2 - x_3 + 8x_4 & = & 15 \end{array}\right\} \tag{3.6}$$

Consideramos $\mathbf{x}^{(0)} = (0, 0, 0, 0)$. La iteraciones del método son:

$$x_1^{(k+1)} = \frac{1}{10}\left(6 + x_2^{(k)} - 2x_3^{(k)}\right)$$

$$x_2^{(k+1)} = \frac{1}{11}\left(25 + x_1^{(k)} + x_3^{(k)} - 3x_4^{(k)}\right)$$

[11] Introducido por el matemático alemán Carl Gustav Jakob Jacobi en 1845.

$$x_3^{(k+1)} = \frac{1}{10}\left(-11 - 2x_1^{(k)} + x_2^{(k)} + x_4^{(k)}\right)$$

$$x_4^{(k+1)} = \frac{1}{8}\left(15 - 3x_2^{(k)} + x_3^{(k)}\right)$$

Así, para calcular $\mathbf{x}^{(1)}$ sustituimos los valores de las componentes de $\mathbf{x}^{(0)}$ en el esquema anterior, obteniendo $\mathbf{x}^{(1)} = (0.6, 2.2727, -1.1, 1.875)$. A continuación, repetimos el proceso sustituyendo los valores obtenidos de $\mathbf{x}^{(1)}$ para calcular $\mathbf{x}^{(2)}$, etc.

$$\mathbf{x}^{(2)} = (1.0472, 1.7159, -0.8052, 0.8852),$$

$$\mathbf{x}^{(3)} = (0.9326, 2.0533, -1.0493, 1.1308),$$

$$\vdots$$

Al cabo de diez iteraciones se tiene que $\mathbf{x}^{(10)} = (1.0001, 1.9997, -0.9998, 0.9997)$. Nótese que la solución del sistema es $(1, 2, -1, 1)$, lo que nos confirma que el método está convergiendo.

### Método de Gauss-Seidel[12]

Si miramos el método de Jacobi con atención podemos observar que la construcción de cada componente de $\mathbf{x}^{(k+1)}$ usa sólo la información obtenida de $\mathbf{x}^{(k)}$; sin embargo, a la hora de calcular $x_i^{(k+1)}$ ya conocemos los valores de $x_j^{(k+1)}$ para $j < i$. El método de Gauss-Seidel simplemente aprovecha este hecho. Así, al igual que antes, comenzaríamos con una aproximación inicial dada $\mathbf{x}^{(0)}$, y en cada iteración, conocida la aproximación $\mathbf{x}^{(k)}$ calculamos $\mathbf{x}^{(k+1)}$ mediante:

$$\begin{aligned}
x_1^{(k+1)} &= \frac{1}{a_{11}}\left(b_1 - a_{12}x_2^{(k)} - \cdots - a_{1n}x_n^{(k)}\right)\\
x_2^{(k+1)} &= \frac{1}{a_{22}}\left(b_2 - a_{21}x_1^{(k+1)} - \cdots - a_{2n}x_n^{(k)}\right)\\
&\vdots\\
x_i^{(k+1)} &= \frac{1}{a_{ii}}\Big(b_i - a_{i1}x_1^{(k+1)} - \cdots - a_{i,i-1}x_{i-1}^{(k+1)} - a_{i,i+1}x_{i+1}^{(k)}\\
&\qquad - \cdots - a_{in}x_n^{(k)}\Big)
\end{aligned}$$

[12] Introducido por el alemán Philipp Ludwig von Seidel en 1874.

$$\vdots$$

$$x_n^{(k+1)} = \frac{1}{a_{nn}} \left( b_n - a_{n1} x_1^{(k+1)} - a_{n2} x_2^{(k+1)} - \cdots - a_{n,n-1} x_{n-1}^{(k+1)} \right)$$

Es de esperar que el método de Gauss-Seidel converja más rápidamente que el método de Jacobi, aunque no siempre es así. Lo que sí se puede probar es que converge, al menos, igual de rápido que el de Jacobi.

**Ejemplo 3.13**

Aplicamos el método de Gauss-Seidel al sistema (3.6). Iniciando desde el mismo valor $\mathbf{x}^{(0)} = (0, 0, 0, 0)$, la iteración a realizar es

$$\begin{aligned}
x_1^{(k+1)} &= \frac{1}{10} \left( 6 + x_2^{(k)} - 2x_3^{(k)} \right) \\
x_2^{(k+1)} &= \frac{1}{11} \left( 25 + x_1^{(k+1)} + x_3^{(k)} - 3x_4^{(k)} \right) \\
x_3^{(k+1)} &= \frac{1}{10} \left( -11 - 2x_1^{(k+1)} + x_2^{(k+1)} + x_4^{(k)} \right) \\
x_4^{(k+1)} &= \frac{1}{8} \left( 15 - 3x_2^{(k+1)} + x_3^{(k+1)} \right)
\end{aligned}$$

Al igual que antes, para obtener $x_1^{(1)}$ sustituimos los valores de las componentes de $\mathbf{x}^{(0)}$ de las que partimos; pero para calcular $x_2^{(1)}$, ya conocemos la componente $x_1^{(1)}$, de manera que la usamos en lugar de $x_1^{(0)}$. Así, procedemos con las demás componentes en cada iteración, obteniéndose

$$\begin{aligned}
\mathbf{x}^{(1)} &= (0.6, 2.3272, -0.9872, 0.8788), \\
\mathbf{x}^{(2)} &= (1.0301, 2.0369, -1.0144, 0.9843), \\
\mathbf{x}^{(3)} &= (1.0065, 2.0035, -1.0025, 0.9983), \\
\mathbf{x}^{(4)} &= (1.0008, 2.0002, -1.0003, 0.9998)
\end{aligned}$$

Vemos que en este caso, con la mitad de iteraciones obtenemos una precisión similar a la obtenida con el método de Jacobi.

Por otra parte, es importante resaltar que la convergencia de ambos métodos no está asegurada. Se pueden dar condiciones suficientes para que ambos métodos converjan. Una de las más habituales es la *diagonal dominancia*.

**Definición 3.5**

Se dice que una matriz $A \in \mathcal{M}_n(\mathbb{K})$ es diagonal dominante (por filas) si

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|, \quad \forall i$$

Es decir, en cada fila, el elemento diagonal en módulo (o valor absoluto) es mayor que la suma de los módulos del resto de elementos de esa fila. El mismo concepto se puede aplicar por columnas. Se puede probar que si una matriz es diagonal dominante, los métodos de Jacobi y Gauss-Seidel convergen.

## 3.4 CÁLCULO CON PYTHON

Como hemos comentado con anterioridad, la necesidad de resolver sistemas de ecuaciones lineales de gran tamaño dio lugar al Álgebra Lineal Numérica de la que hemos descrito unas pinceladas en la sección anterior. Python, como lenguaje de programación que es, es un entorno adecuado para programar los diversos algoritmos que hemos visto, sin embargo, no queremos dedicar nuestra atención en este texto al aprendizaje de la programación (aunque se verán algunos ejemplos) sino al uso de algunas de las herramientas que Python tiene incorporadas para resolver sistemas de ecuaciones.

La solución de sistemas de ecuaciones desde NumPy es sencilla; basta invocar la función `linalg.solve` con la matriz del sistema y el segundo miembro: por ejemplo, el código

```
>>> from numpy import matrix,linalg
>>> a=matrix('2. 1 1; 1 2 1 ; 1 1 2 ')
>>> b=matrix('1; 2; 3')
>>> linalg.solve(a,b)
matrix([[-0.5],
        [ 0.5],
        [ 1.5]])
```

resuelve el sistema

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

cuya solución es $x_1 = -0.5$, $x_2 = 0.5$ y $x_3 = 1.5$.

Si la matriz es no cuadrada o singular, obtendremos un error.

Por su parte, SymPy nos ofrece más sofisticación, permitiendo resolver sistemas indeterminados, aunque su utilidad es escasa si el sistema es grande. Por ejemplo,

```
>>> from sympy import Matrix,solve_linear_system
>>> from sympy.abc import x,y,z
>>> A=Matrix([[-2,1,1,0],[1,-2,1,0],[1,1,-2,0]])
>>> solve_linear_system(A,x,y,z)
{x: z, y: z}
```

corresponde a la solución de

$$\left.\begin{array}{rcl} -2x+y+z & = & 0 \\ x-2y+z & = & 0 \\ x+y-2z & = & 0 \end{array}\right\}$$

que es un sistema compatible indeterminado con solución: $(\alpha, \alpha, \alpha)$.

Vemos que es necesario pasar como argumento la matriz ampliada del sistema y los nombres de las variables que reciben la solución.

Otra opción para resolver un sistema pasa por el uso del módulo `rref()` de SymPy que ya vimos en el tema anterior. Este módulo permite hacer una reducción tipo Gauss de una matriz, y por tanto nos está proporcionando de forma indirecta la solución de un sistema.

```
>>> from sympy import Matrix
>>> a=Matrix([[2,1,1,1],[1,2,1,2],[1,1,2,3]])
>>> a.rref()
([1, 0, 0, -1/2]
[0, 1, 0,  1/2]
[0, 0, 1,  3/2], [0, 1, 2])
```

corresponde a

$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 3 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} \end{array}\right)$$

Si miramos la última columna de la matriz resultante, ésta es precisamente la solución del sistema. Como ya comentamos, `rref()` realiza una reducción tipo Gauss, hasta obtener una matriz diagonal, con unos en la diagonal (en este caso, la matriz identidad). Luego el segundo miembro del sistema (la última columna) resulta la solución (véase el ejemplo 2.10).

La factorización LU de un matriz también es posible, aunque no está definida en NumPy, sino en un módulo habitual para el cálculo científico conocido como SciPy.[13]

[13] `www.scipy.org`

```
1  >>> from numpy import matrix
2  >>> from scipy import linalg
3  >>> a=matrix('2. 1 1; 1 2 1 ; 1 1 2 ')
4  >>> p,l,u=linalg.lu(a)
5  >>> p
6  array([[ 1.,  0.,  0.],
7         [ 0.,  1.,  0.],
8         [ 0.,  0.,  1.]])
9  >>> l
10 array([[ 1.        ,  0.        ,  0.        ],
11        [ 0.5       ,  1.        ,  0.        ],
12        [ 0.5       ,  0.33333333,  1.        ]])
13 >>> u
14 array([[ 2.        ,  1.        ,  1.        ],
15        [ 0.        ,  1.5       ,  0.5       ],
16        [ 0.        ,  0.        ,  1.33333333]])
```

es decir,

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{3} \end{pmatrix}$$

Obsérvese el tipo de llamada (línea 4): la función `lu` proporciona tres resultados (las matrices $P$, $L$ y $U$, por ese orden), que son almacenadas en las variables `p`, `l` y `u`, que, como ya hemos visto, corresponden a la matriz de permutación $P$ y las matrices $L$ y $U$ que verifican $PA = LU$. Nótese que en este caso $P$ es la matriz identidad.

Si por el contrario escogemos el módulo SymPy:

```
>>> from sympy import Matrix
>>> A=Matrix([[0,3,2],[1,0,-3],[-2,1,-1]])
>>> l,u,p=A.LUdecomposition()
>>> l
[ 1,   0, 0]
[ 0,   1, 0]
[-2, 1/3, 1]
>>> u
[1, 0,     -3]
[0, 3,      2]
[0, 0, -23/3]
>>> p
[[1, 0]]
```

vemos que la llamada difiere con respecto a la descomposición LU de SciPy. En este caso se obtienen $L$, $U$ y $P$, en ese orden, teniendo en cuenta que $P$ no viene dada en forma de matriz, sino a través de los índices de las filas que han

permutado (en este caso la primera con la segunda). En definitiva:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 3 & 2 \\ 1 & 0 & -3 \\ -2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & \frac{1}{3} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -3 \\ 0 & 3 & 2 \\ 0 & 0 & -\frac{23}{3} \end{pmatrix}$$

Por otra parte, NumPy también permite el cálculo del número de condición de una matriz:

```
>>> from numpy import matrix,linalg
>>> a=matrix('2 3; 2+1e-7 3')
>>> a
matrix([[ 2.        ,  3.        ],
        [ 2.0000001,  3.        ]])
>>> linalg.cond(a)
86666667.760877177
```

esto es,

$$\kappa \begin{pmatrix} 2 & 3 \\ 2+10^{-7} & 3 \end{pmatrix} \approx 86666667.76$$

indicándonos el mal condicionamiento de la matriz del ejemplo 3.11.

Finalmente, veamos un pequeño programa para realizar las iteraciones del método de Jacobi mostrado en la sección anterior. Con objeto de implementar el código de forma eficiente en Python, hay que evitar en la medida de lo posible la aparición de bucles, pues éstos ralentizan enormemente la ejecución de los programas. Para evitar bucles la mejor opción es "vectorizar" las operaciones que involucran un arreglo o matriz, pues éstas se computan mucho más rápidamente.

Así pues, vamos a reinterpretar las operaciones del método de Jacobi del siguiente modo. Si observamos el segundo miembro de (3.5) vemos que corresponde al producto de los inversos de los elementos diagonales de la matriz $A$, por un paréntesis que corresponde a la diferencia entre $\mathbf{b}$ y el producto de la matriz $A$, sin los elementos diagonales, por $\mathbf{x}$.

Veámoslo del siguiente modo: sea $D$ la matriz diagonal, cuya diagonal es la misma que la de la matriz $A$, y sea $\tilde{A} = A - D$. El sistema $A\mathbf{x} = \mathbf{b}$ se puede escribir como

$$(\tilde{A} + D)\mathbf{x} = \mathbf{b} \Longrightarrow D\mathbf{x} = \mathbf{b} - \tilde{A}\mathbf{x} \Longrightarrow \mathbf{x} = D^{-1}\mathbf{b} - D^{-1}\tilde{A}\mathbf{x}$$

Observando el último término, vemos que corresponde exactamente con el segundo miembro de (3.5). Por tanto, las iteraciones del método de Jacobi ahora son:

$$\mathbf{x}^{(k+1)} = D^{-1}\mathbf{b} - D^{-1}\tilde{A}\mathbf{x}^{(k)} \tag{3.7}$$

Definimos entonces una función en Python que, introducidas la matriz del sistema, el segundo miembro, una aproximación inicial y el número de iteraciones a realizar, devuelve el valor de la última iteración calculada:

```
1  from numpy import diagflat,zeros,linalg
2  from sys import exit
3  def jacobi(a,b,x0=None,iter=10):
4  
5      if a.shape[0]!=a.shape[1]:
6          exit('la matriz no es cuadrada')
7      elif (b.shape[1]!=1) or (b.shape[0]!=a.shape[0]):
8          exit('el segundo miembro no es correcto')
9      if x0 is None:
10         x0=zeros((a.shape[0],1))
11     elif(b.shape!=x0.shape):
12         exit('dato inicial incorrecto')
13 
14     d=diagflat(a.diagonal())
15     if (linalg.det(d)==0):
16         exit('hay elementos diagonales nulos')
17     atd=d.I*(a-d)
18     c=d.I*b
19     for i in range(iter):
20         x=c-atd*x0
21         x0=x
22 
23     return x0
```

Las dos primeras líneas realizan la importación de las funciones necesarias para el programa. Destacamos en este código la aparición de la instrucción `exit()` del módulo `sys`, que permite abandonar la ejecución del mismo. Dicha instrucción es usada entre las líneas 5–8 en caso de que se cumpla alguna de las condiciones `if` expuestas. Estas condiciones comprueban que los argumentos de entrada son correctos: en concreto, se analiza si el dato `a` es una matriz cuadrada y si el dato `b` es una matriz columna con la misma dimensión que `a`. Si no se cumple alguna de estas condiciones la ejecución del programa se detiene con la impresión del mensaje correspondiente.

Otra novedad del código es la aparición de argumentos por defecto en la llamada a la función (línea 3). Puesto que es habitual que la aproximación inicial sea una matriz de ceros, podemos omitir la aproximación inicial en la llamada a la función, y en ese caso, el programa crea una aproximación inicial nula (líneas 9–10). En caso de que sí se haya introducido una aproximación inicial se comprueba que ésta tiene las dimensiones adecuadas.

La línea 14 crea la matriz `d` correspondiente a la diagonal de la matriz del sistema. Para ello se ha usado el atributo `diagonal()` con el que se obtiene una matriz fila correspondiente a la diagonal de la matriz `a`, y con la función `diagflat` se construye una matriz diagonal, cuyos elementos diagonales son los

de la matriz fila usada como argumento.

A continuación, en las líneas 17 y 18 se construyen los elementos para la iteración, según hemos visto en (3.7). Nótese también que comprobamos que la matriz `d` es invertible calculando su determinante.

Llegados a este punto es interesante notar que, hasta el momento, no se han realizado más que pequeñas comprobaciones y cálculos preparativos. Las iteraciones del método de Jacobi se llevan a cabo en las líneas 19–21. Observar también que en cada iteración, actualizamos el valor de la aproximación, para poder calcular la siguiente.

Para poder usar esta función precisamos los datos del sistema y cargarla previamente. Suponiendo que el código anterior es guardado en un archivo de nombre `jaco.py`, el siguiente código muestra cómo resolver el ejemplo 3.13

```
>>> from numpy import matrix
>>> from jaco import jacobi
>>> a=matrix('10 -1 2 0; -1 11 -1 3; 2 -1 10 -1; 0 3 -1 8')
>>> b=matrix('6; 25; -11; 15')
>>> jacobi(a,b)
matrix([[ 1.0001186 ],
        [ 1.99976795],
        [-0.99982814],
        [ 0.99978598]])
```

En el ejercicio 23 proponemos al lector la elaboración de un código similar para el método de Gauss-Seidel.

## 3 5 APLICACIÓN: RESOLUCIÓN DE ECUACIONES DIFERENCIALES

Como hemos comentado anteriormente, los sistemas de ecuaciones lineales aparecen frecuentemente en muchas aplicaciones de las Matemáticas y muy especialmente en la resolución de ecuaciones diferenciales. Aunque este es un tema que se escapa de los conocimientos típicos de los lectores a los que va dirigido este texto, creemos importante mostrar cómo surgen sistemas de ecuaciones lineales con un gran número de ecuaciones e incógnitas a través de un ejemplo sencillo.

No es nuestra intención entrar en profundidad en demasiados aspectos acerca de las ecuaciones diferenciales; basta decir que se trata de ecuaciones en las que la incógnita es una función (y no simplemente un conjunto de valores), y que dicha función se haya afectada bajo alguna operación que involucra la derivación. Un ejemplo típico de ecuación diferencial puede ser:

$$u''(x) = 0, \quad x \in (0,1)$$

Es decir, se trataría de buscar una función $u$ definida en el intervalo $(0,1)$ de forma que su segunda derivada sea cero, en cualquiera de sus puntos. Con unos

conocimientos básicos de derivación podemos ver que cualquier función de la forma $u(x) = ax + b$, para cualesquiera valores $a$ y $b$, es una solución de dicha ecuación. Con objeto de poder encontrar una única solución de la ecuación, es habitual completarla con alguna condición adicional, como puede ser el valor de la función en los extremos del intervalo:

$$u(0) = 1, \qquad u(1) = 3$$

Con estas condiciones podemos determinar los valores de $a$ y $b$ resultando que $a = 2$ y $b = 1$. Así pues, la función que cumple la ecuación diferencial junto con las condiciones en los extremos es la recta $u(x) = 2x + 1$.

Esta ecuación diferencial junto con las condiciones adicionales, denominadas *condiciones de contorno*, modelan, por ejemplo, la distribución de temperatura de una varilla unidimensional de longitud 1 que no está sometida a ninguna fuente de calor, y en cuyos extremos existe una temperatura dada por los valores de las condiciones de contorno. Si añadimos una fuente de calor sobre la varilla, medida a través de una función $f(x)$, que se supone conocida, el modelo matemático que nos proporciona cómo se distribuye el calor en la varilla viene dado por:

$$\begin{gathered} -u''(x) = f(x), \quad x \in (0,1) \\ u(0) = \alpha, \quad u(1) = \beta \end{gathered} \tag{3.8}$$

donde $\alpha$ y $\beta$ son las condiciones de contorno, que se suponen dadas.

Para ciertas funciones $f$ es posible calcular analíticamente la solución de esta ecuación diferencial (integrando dos veces $f$), pero en general puede ser difícil, o incluso imposible, encontrar una expresión explícita para $u$. En estos casos es necesario acudir al Análisis Numérico. En primer lugar hemos de rebajar nuestra pretensión de encontrar la solución como una función, con sus infinitos valores en el intervalo $(0, 1)$, y buscar en su lugar un número finito de valores de la misma; esto es lo que se conoce como *discretización*.

La discretización pasa por considerar un conjunto finito de puntos en el intervalo de partida $(0, 1)$. Lo más habitual consiste en tomar una serie de puntos uniformemente espaciados, que denominamos *nodos*. Para ello tomamos $n \in \mathbb{N}$ y $h = \frac{1}{n+1}$, el denominado *paso de discretización*, y creamos los $n + 2$ puntos siguientes del intervalo $(0, 1)$:

$$x_i = ih, \quad 0 \le i \le n + 1$$

Observemos que $x_0 = 0 < x_1 < x_2 < \cdots < x_n < x_{n+1} = 1$. La intención es encontrar los valores de la solución de (3.8) en estos nodos. Por ejemplo, ya sabemos que $u(x_0) = \alpha$ y $u(x_{n+1}) = \beta$, y por tanto nos quedan por encontrar $u(x_i)$, $1 \le i \le n$, es decir, $n$ valores que serán nuestras incógnitas. Por otra parte, parece evidente pensar que cuantos más puntos consideremos en la discretización mejor será la aproximación de la solución que obtengamos.

El siguiente paso consiste en escribir la ecuación $-u''(x) = f(x)$ en función de esos valores. La cuestión es cómo evaluar $u''$. Es conocido que el valor de

la derivada en un punto $x$ se puede aproximar por cocientes incrementales del tipo:

$$u'(x) \approx \frac{u(x+t)-u(x)}{t} \approx \frac{u(x-t)-u(x)}{t} \approx \frac{u(x+t)-u(x-t)}{2t}$$

y de forma similar se puede obtener una aproximación de la derivada segunda:

$$u''(x) \approx \frac{u(x+t)-2u(x)+u(x-t)}{t^2}$$

Estas expresiones son más precisas cuanto más pequeño es $t$. Si ponemos $t = h$ en la última expresión y evaluamos en un nodo $x_i$,

$$u''(x_i) \approx \frac{u(x_{i+1})-2u(x_i)+u(x_{i-1})}{h^2}$$

debido a la definición que hemos tomado de los nodos. Con objeto de simplificar la notación pondremos $u(x_i) = u_i$, y de forma similar $f(x_i) = f_i$.

Lo que hacemos a continuación es sencillo: sustituimos la ecuación en $x$ por la evaluación de la misma en cada uno de los nodos obteniendo el siguiente *problema discreto*:

$$-\frac{u_{i+1}-2u_i+u_{i-1}}{h^2} = f_i, \quad 1 \le i \le n \tag{3.9}$$

Dado que $u_0 = \alpha$, $u_{n+1} = \beta$ y $f_i$, $1 \le i \le n$ son valores conocidos, estamos ante un sistema lineal de $n$ ecuaciones con $n$ incógnitas. Vemos por tanto cómo una discretización de una ecuación diferencial desemboca de forma bastante natural en la resolución de un sistema lineal, que será tanto más grande, cuanto mayor sea el valor de $n$, lo que a su vez conducirá, en principio, a una mejor aproximación de los valores de la solución.

En este punto es preciso hacer alguna puntualización: una discretización de una ecuación diferencial cualquiera no siempre conduce a un sistema lineal, y no siempre la solución de un problema discreto proporciona una buena aproximación del problema original; el Análisis Numérico requiere prestar atención a determinados aspectos que quedan fuera de nuestro interés en este texto. En lo que respecta a este ejemplo, sí es posible probar que, para un buen número de funciones $f$, la discretización realizada es coherente con nuestra intuición inicial; cuanto mayor es $n$, mejor es la aproximación obtenida a la solución de la ecuación diferencial.

Prestemos ahora un poco más de atención al sistema (3.9): si escribimos una a una sus ecuaciones resulta,

$$\begin{array}{ccc}
-u_0+2u_1-u_2 & = & h^2 f_1 \\
-u_1+2u_2-u_3 & = & h^2 f_2 \\
\vdots & \vdots & \vdots \\
-u_{n-2}+2u_{n-1}-u_n & = & h^2 f_{n-1} \\
-u_{n-1}+2u_n-u_{n+1} & = & h^2 f_n
\end{array}$$

que escribimos en forma matricial como

$$\begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} = h^2 \begin{pmatrix} f_1 + \alpha \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n + \beta \end{pmatrix} \tag{3.10}$$

La matriz del sistema tiene propiedades interesantes. Es lo que se conoce como una matriz vacía, esto es, con muchos ceros, que además tiene una estructura especial. Además de ser simétrica, es una matriz *banda*, y más concretamente *tridiagonal*. Este tipo de matrices son frecuentes en la aproximación numérica de ecuaciones diferenciales y ecuaciones en derivadas parciales, y dadas sus especiales características, reciben un tratamiento numérico especial en el que no entraremos.

Para terminar, veamos cómo resolver un ejemplo con Python.

Resolvamos numéricamente la ecuación diferencial

$$\begin{aligned} -u''(x) &= 4\pi^2 \operatorname{sen}(2\pi x), \qquad x \in (0,1) \\ u(0) &= u(1) = 0 \end{aligned}$$

Vamos a crear una función (al estilo de la función `mandelplot` del tema 1) de tal forma que, dado $n$ el número de nodos, la función $f$ y los valores de contorno $\alpha$ y $\beta$, proporcione la solución de (3.8).

```
#! /usr/bin/python

from numpy import pi,linspace,sin,diag,ones,linalg,append
from matplotlib.pyplot import plot,show,grid,axis

def calor(n, f, alfa, beta):

    x=linspace(0.,1.,n+2)
    h=1./(n+1)
    fx=h**2*f(x[1:n+1])
    fx[0]+=alfa
    fx[n-1]+=beta
    a=diag(2.*ones(n)) + diag(-1.*ones(n-1),1) + diag(-1.*
        ones(n-1),-1)

    u=linalg.solve(a,fx)
    u=append(alfa,u)
    u=append(u,beta)

    return u,x
```

```
20
21
22 f= lambda x:  4*pi**2*sin(2*pi*x)
23 n=40
24 alfa=0.
25 beta=0.
26
27 u,x=calor(n,f,alfa,beta)
28
29 g=lambda x: sin(2*pi*x)
30
31 plot(x,u,'k-',marker='*')
32 plot(x,g(x),'r--')
33 grid(True)
34 axis([0,1,-1.2,1.2])
35 show()
```

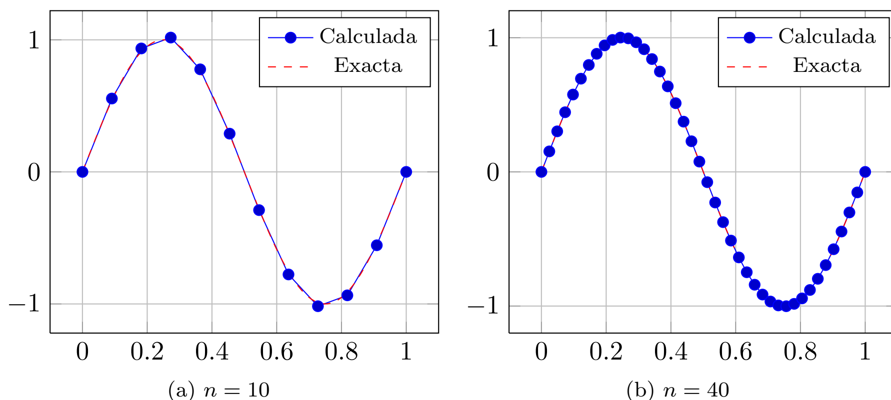
Analicemos con detenimiento el código. Tras la habitual importación de funciones de los módulos apropiados (nótese por ejemplo la importación de la constante `pi` o la función seno desde NumPy), definimos la función `calor`, entre las líneas 6–19, que es la que se encarga de resolver la ecuación diferencial.

En primer lugar creamos los nodos en el intervalo $(0,1)$ a través de `linspace`. Esta función crea un arreglo formado por $n+2$ valores entre 0 y 1. A continuación calculamos el paso de discretización y en la línea 10 construimos el segundo miembro de (3.9). Nótese que evaluamos la función dato del problema en los nodos de un solo golpe, sin necesidad de usar un bucle, excluyendo los valores de los extremos pues no son necesarios. En las líneas 11 y 12 rectificamos el primer y el último valor de este arreglo según indica el segundo miembro de (3.10), y en la línea 13 construimos la matriz del sistema usando una combinación de la función `diag`.

Veamos esta construcción con más atención: `diag(2.*ones(n))` crea una matriz diagonal cuya diagonal principal está formada por `2.*ones(n)`, que corresponde a un arreglo de dimensión $n$ formado por unos (esto es lo que hace `ones(n)`), y que está multiplicado por 2. Es decir, es una matriz diagonal con 2 en la diagonal. A esta matriz se le suma `diag(-1.*ones(n-1),1)`; está claro qué es `-1.*ones(n-1)`, mientras que el segundo argumento de `diag` *desplaza* una diagonal hacia arriba el lugar donde se sitúan, en este caso, los $-1$. Algo similar hace `diag(-1.*ones(n-1),-1)`, pero en lugar de desplazar hacia arriba, lo hace hacia abajo. Ahora es inmediato ver que la suma de estos tres términos es la matriz de (3.10).

Finalmente, la línea 15 resuelve el sistema, mientras que las líneas 16 y 17 incorporan los valores en los extremos a la solución obtenida. En la línea 19 se produce la salida de la función, en la que devolvemos tanto la solución `u` como los nodos `x`, con objeto de poder dibujarlos si es necesario.

La ejecución de la función se lleva a cabo entre las líneas 22 y 27. En la línea 22 definimos la función dato del problema, en este caso, $f(x) = 4\pi^2 \operatorname{sen}(2\pi x)$.




(a) $n = 10$ (b) $n = 40$

Figura 3.1: Solución de la ecuación del calor

Si bien se podría haber creado dicha función mediante

```
def f(x):
    return 4*pi**2*sin(2*pi*x)
```

la función `lambda` simplifica esta definición. Entre las líneas 23 y 25 fijamos el resto de parámetros de nuestro problema, y en la línea 27 llamamos a la función `calor` para resolver el problema.

Con objeto de comprobar la precisión del método empleado para resolver la ecuación diferencial, hemos comparado la solución obtenida con la solución exacta, que en este caso conocemos y es $g(x) = \operatorname{sen}(2\pi x)$. Para ello hemos dibujado ambas soluciones con la función `plot` del módulo `matplotlib`. En la figura 3.1 se puede ver el resultado para un par de valores de $n$ distintos.

3 6

## EJERCICIOS

### Ejercicios de repaso

**E.1** Resolver los siguientes sistemas mediante el método de eliminación de Gauss:

$$\text{(a)} \quad \left.\begin{array}{rcl} x_1 + x_2 + x_3 &=& 6 \\ x_1 + 2x_2 + 2x_3 &=& 9 \\ x_1 + 2x_2 + 3x_3 &=& 10 \end{array}\right\} \qquad \text{(b)} \quad \left.\begin{array}{rcl} x_1 + 2x_2 &=& 3 \\ -x_2 + x_3 &=& 1 \\ 3x_1 + x_2 + 5x_3 &=& 0 \end{array}\right\}$$

**E.2** Usar la Regla de Cramer para resolver los sistemas:

$$\text{(a)} \left.\begin{array}{rcl} 2x-y-z &=& 4 \\ 3x+4y-2z &=& 11 \\ 3x-2y+4z &=& 11 \end{array}\right\} \qquad \text{(b)} \left.\begin{array}{rcl} x+y+2z &=& -1 \\ 2x-y+2z &=& -4 \\ 4x+y+4z &=& -2 \end{array}\right\}$$

**E.3** Un sistema de ecuaciones de matriz $A \in \mathcal{M}_{4\times 8}$ tiene 5 grados de libertad. Calcula rango$(A)$.

**E.4** Utilizar el teorema de Rouché-Frobenius para realizar un estudio del número de soluciones de los siguientes sistemas y resolverlos:

$$\text{(a)} \left.\begin{array}{rcl} 3x_1+2x_2+x_4-x_5 &=& 0 \\ x_1+2x_3+x_5 &=& 3 \end{array}\right\} \qquad \text{(b)} \left.\begin{array}{rcl} 2x_1+x_2 &=& 5 \\ x_1-x_2 &=& 1 \\ x_1+2x_2 &=& 0 \end{array}\right\}$$

$$\text{(c)} \left.\begin{array}{rcl} x_1+2x_2-x_3 &=& 1 \\ x_2+x_3 &=& 2 \\ x_1+3x_2 &=& 3 \end{array}\right\} \qquad \text{(d)} \left.\begin{array}{rcl} x_1+x_3-x_4 &=& 5 \\ x_2+x_3+x_4 &=& 2 \end{array}\right\}$$

**E.5** En los siguientes apartados se da un número $x$ y una aproximación $x^*$. Encontrar los errores absoluto y relativo en cada caso.

(a) $x = 5$; $x^* = 0.49 \times 10^1$.

(b) $x = 0.704645$; $x^* = 0.70466 \times 10^0$.

**E.6** Resuelve el siguiente sistema mediante eliminación gaussiana con y sin pivoteo parcial, usando aritmética de coma flotante con cuatro cifras significativas. Después encuentra la solución exacta y calcula los errores relativos de los valores calculados:

$$\left.\begin{array}{rcr} 2x_1+\frac{2}{3}x_2+\frac{1}{3}x_3 &=& 1 \\ x_1+2x_2-x_3 &=& 0 \\ 6x_1+2x_2+2x_3 &=& -2 \end{array}\right\}$$

**E.7** Hallar la factorización LU de las siguientes matrices

$$\text{(a)} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 3 \\ 1 & 3 & 2 & 1 \\ 2 & 2 & 1 & 1 \end{pmatrix} \qquad \text{(b)} \begin{pmatrix} 1 & -1 & 1 & 0 \\ -1 & 2 & -1 & 2 \\ 1 & -1 & 5 & 2 \\ 0 & 2 & 2 & 4 \end{pmatrix}$$

$$\text{(c)} \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

E.8 Demuestre que el sistema

$$\left.\begin{aligned} x_1 + x_2 &= 50 \\ x_1 + 1.026x_2 &= 20 \end{aligned}\right\}$$

está mal condicionado si se usa aritmética de coma flotante con dos cifras significativas. ¿Cuál es el error relativo cometido si lo comparamos con la solución exacta?

E.9 Obtener las tres primeras iteraciones del método de Jacobi para los siguientes sistemas usando $\mathbf{x}^{(0)} = (0, 0, 0)$:

$$\text{(a)} \quad \left.\begin{array}{rcl} 3x_1 - x_2 + x_3 & = & 1 \\ 3x_1 + 6x_2 + 2x_3 & = & 0 \\ 3x_1 + 3x_2 + 7x_3 & = & 4 \end{array}\right\} \qquad \text{(b)} \quad \left.\begin{array}{rcl} 10x_1 - x_2 & = & 9 \\ -x_1 + 10x_2 - 2x_3 & = & 7 \\ -2x_2 + 10x_3 & = & 6 \end{array}\right\}$$

E.10 Repetir el ejercicio 9 empleando el método de Gauss-Seidel.

## Problemas

E.11 ¿Puede existir una parábola que pase por los puntos $(0, 1)$, $(1, 3)$, $(2, 15)$ y $(3, 37)$?

E.12 ¿Para qué valores de $m$ el sistema

$$\left.\begin{array}{rcl} 2x - y + z & = & 0 \\ x + my - z & = & 0 \\ x + y + z & = & 0 \end{array}\right\}$$

tiene soluciones no triviales?

E.13 Estudiar el sistema:

$$\left.\begin{array}{rcl} 4x + 2y + z & = & \lambda x \\ 2x + 4y + 2z & = & \lambda y \\ 2x + 4y + 8z & = & \lambda z \end{array}\right\}$$

según los valores del parámetro $\lambda$.

E.14 Discutir la compatibilidad de los sistemas siguientes según los valores de los parámetros $m$ y $n$.

$$\text{(a)} \quad \left.\begin{array}{rcl} 3x - 7y & = & m \\ x + y & = & n \\ 5x - 13y & = & 5m - 2n \\ x + 2y & = & m + n - 1 \end{array}\right\} \qquad \text{(b)} \quad \left.\begin{array}{rcl} x - my + z & = & 0 \\ x + y - z & = & 0 \\ nx - 2y - 5z & = & 0 \\ 2x + y + z & = & 0 \end{array}\right\}$$

**E.15** Construir un sistema lineal de tres ecuaciones y cuatro incógnitas tal que su solución sea $(1-2\alpha-3\beta,\alpha,1+2\beta,\beta)$.

**E.16** Un sistema con tres incógnitas y un grado de libertad tiene como soluciones las ternas $(1,0,1)$ y $(2,-1,-1)$. Encuentra todas las soluciones del sistema.

**E.17** Encuentra la matriz cuadrada $P$ tal que al multiplicar a la derecha por $(x,y,z)^T$ da $(y,z,x)^T$. ¿Cuál es su inversa?

**E.18** Resolver los siguientes sistemas usando la factorización LU:

$$\text{(a)}\quad \left.\begin{array}{rcl} x-y+z &=& 4\\ -x+2y-z+2t &=& -3\\ x-y+5z+2t &=& 16\\ 2y+2z+6t &=& 8 \end{array}\right\} \qquad \text{(b)}\quad \left.\begin{array}{rcl} x+y &=& 5\\ -y+5z &=& 2\\ x+2y-z &=& 7 \end{array}\right\}$$

**E.19** Comprobar que el sistema

$$\begin{aligned} x_1-x_2-x_3-x_4-x_5&=0\\ x_2-x_3-x_4-x_5&=0\\ x_3-x_4-x_5&=0\\ x_4-x_5&=0\\ x_5&=1 \end{aligned}$$

está mal condicionado, si se considera el sistema perturbado

$$\begin{aligned} x_1-x_2-x_3-x_4-x_5&=0\\ -\tfrac{1}{15}x_1+x_2-x_3-x_4-x_5&=0\\ -\tfrac{1}{15}x_1+x_3-x_4-x_5&=0\\ -\tfrac{1}{15}x_1+x_4-x_5&=0\\ -\tfrac{1}{15}x_1+x_5&=1 \end{aligned}$$

### Ejercicios teóricos

**E.20** Probar la Proposición 3.1

**E.21** Decidir si las siguientes afirmaciones son verdaderas o falsas:

(a) Un sistema compatible indeterminado puede tener el mismo número de ecuaciones e incógnitas.

(b) Un sistema de $n+1$ ecuaciones con $n$ incógnitas tal que el rango de su matriz ampliada sea $n+1$ puede ser indeterminado.

(c) Un sistema compatible determinado puede tener más incógnitas que ecuaciones.

(d) Un sistema compatible determinado puede tener más ecuaciones que incógnitas.

E.22 Construir un sistema lineal de matriz no nula con más incógnitas que ecuaciones que no tenga solución. ¿Cuál es el menor número de ecuaciones e incógnitas que puede tener?

*E.23 Considérese el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ con $A \in \mathcal{M}_n(\mathbb{K})$, $\mathbf{x}, \mathbf{b} \in \mathcal{M}_{n\times 1}(\mathbb{K})$. Compruébese que las iteraciones del método de Gauss-Seidel equivalen a realizar las iteraciones

$$\mathbf{x}^{(k+1)} = \mathbf{c} - M\mathbf{x}^{(k)}$$

con $\mathbf{c} = (D-L)^{-1}\mathbf{b}$ y $M = (D-L)^{-1}U$, donde $D$, $L$ y $U$ son, respectivamente, una matriz diagonal, cuya diagonal corresponde a la diagonal de $A$, una matriz triangular inferior que corresponde a la parte triangular inferior de la matriz $A$, y una matriz triangular superior que coincide con la parte triangular superior de $A$, de tal modo que $A = D + L + U$.

## Ejercicios adicionales

E.24 Elaborar un programa en Python para resolver un sistema de ecuaciones usando el método de Gauss-Seidel empleando la descomposición que aparece en el ejercicio 23.

E.25 Modifica adecuadamente el código de la página 127 para poder resolver una ecuación diferencial del tipo:

$$\begin{gathered}-u''(x) + u(x) = f(x), \quad x \in (0,1)\\ u(0) = \alpha, \quad u(1) = \beta\end{gathered}$$

# 4 Espacios vectoriales

Los *espacios vectoriales*[1] son probablemente las estructuras matemáticas más comunes que podemos encontrar. Todos los fenómenos calificados como “lineales” en multitud de contextos están vinculados de algún modo a un espacio vectorial, lo que da una idea de su importancia. Por otra parte, son estructuras muy sencillas que entrañan una interesante diversidad de propiedades, algunas de las cuales veremos en este tema.

Posiblemente el lector habrá manejado con anterioridad el concepto de vector, bien sea como elemento geométrico para determinar direcciones, o bien como un objeto que permite representar determinadas magnitudes físicas como la velocidad o la fuerza. En estos casos, el vector se representa como una “flecha” que determina unas características propias como son su módulo, dirección y sentido.[2] Esta representación es útil para “visualizar” ciertos conceptos y propiedades, pero es completamente inadecuada cuando tratamos con otro tipo de objetos que también son vectores, como veremos a lo largo de este tema.

## 4 1 LA ESTRUCTURA DE ESPACIO VECTORIAL

Con objeto de poder tratar cualquier tipo de vector con independencia de su naturaleza, usando simplemente las propiedades intrínsecas que poseen debido a su pertenencia a un espacio vectorial, hemos de hacer un esfuerzo en tratar todos estos conceptos de forma abstracta; la definición de espacio vectorial es una buena muestra de ello. Emplazamos al lector para que trate de entender el significado de la definición a través de los ejemplos que le siguen.

[1] El origen del concepto está vinculado a los trabajos de matemáticos del s. XVII en geometría analítica, matrices y sistemas de ecuaciones lineales, aunque la formulación axiomática actual se debe al matemático italiano Giuseppe Peano a finales del s. XIX, que había estudiado profundamente la obra del alemán Hermann Grassmann de mediados del mismo siglo, en el que de manera no formal se establecen las ideas principales de lo que hoy conocemos como Álgebra Lineal.

[2] El concepto de segmento orientado o *bipoint* se debe al italiano Giusto Bellavitis en el s. XIX, aunque el nombre de vector fue introducido por el irlandés William Rowan Hamilton.

**Definición 4.1**

Un conjunto $V$ se dice que es un *espacio vectorial sobre el cuerpo* $\mathbb{K}$ (o un abreviadamente un $\mathbb{K}$-espacio vectorial) si existen en él las siguientes operaciones:
una operación interna (*suma*):

$$\begin{array}{rccc} +: & V \times V & \longrightarrow & V \\ & (\mathbf{x}, \mathbf{y}) & \longrightarrow & \mathbf{x} + \mathbf{y} \end{array}$$

y una operación externa (*producto por escalar*):

$$\begin{array}{rccc} \cdot: & \mathbb{K} \times V & \longrightarrow & V \\ & (\alpha, \mathbf{x}) & \longrightarrow & \alpha \cdot \mathbf{x} \end{array}$$

verificando:

(I) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(II) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

(III) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(IV) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

(V) $1 \cdot \mathbf{u} = \mathbf{u}$

(VI) $\alpha \cdot (\beta \cdot \mathbf{u}) = (\alpha\beta) \cdot \mathbf{u}$

(VII) $(\alpha + \beta) \cdot \mathbf{u} = \alpha \cdot \mathbf{u} + \beta \cdot \mathbf{u}$

(VIII) $\alpha \cdot (\mathbf{u} + \mathbf{v}) = \alpha \cdot \mathbf{u} + \alpha \cdot \mathbf{v}$

Obsérvese que la definición necesita de la existencia de un determinado cuerpo $\mathbb{K}$ (nosotros usaremos $\mathbb{R}$ ó $\mathbb{C}$) y en función de éste puede variar la estructura del espacio. Si $\mathbb{K} = \mathbb{R}$ se llama espacio vectorial *real* y si $\mathbb{K} = \mathbb{C}$ se dice espacio vectorial *complejo*. Por simplicidad hablaremos simplemente de espacio vectorial (en adelante e.v.) sin hacer mención al cuerpo, siendo el contexto en el que trabajemos el que determine el cuerpo a usar (de forma genérica $\mathbb{K}$).

Como se puede observar, la estructura de e.v. solo precisa de la existencia de un par de operaciones, y de que éstas satisfagan una cuantas propiedades que imaginamos que el lector puede claramente identificar. A través de los ejemplos que siguen a continuación nos daremos cuenta de que prácticamente cualquier conjunto en el que se puedan realizar operaciones numéricas tiene estructura de e.v.

A los elementos de un e.v. se les denomina *vectores*.[3] Los denotaremos por letras en **negrilla**, **u**, **v**,... El punto correspondiente al producto por escalares será habitualmente omitido.

**Ejemplo 4.1**

(I) $\mathbb{R}^n$ es un e.v. (real) con las operaciones siguientes:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$
$$\alpha \cdot (x_1, \dots, x_n) = (\alpha x_1, \dots, \alpha x_n)$$

En particular, $\mathbb{R}$ (para $n = 1$), es un espacio vectorial sobre sí mismo. Sin embargo, si el cuerpo escogido es $\mathbb{C}$, en lugar de $\mathbb{R}$, entonces $\mathbb{R}^n$ pierde la estructura de e.v., pues la operación producto por escalar de un elemento de $\mathbb{R}^n$ con un número complejo no resulta, en general, un elemento de $\mathbb{R}^n$.

(II) $\mathcal{M}_{m\times n}(\mathbb{K})$ es un e.v.sobre $\mathbb{K}$ con las operaciones suma y producto por escalar dadas en la Definición 2.5 (tal y como confirman las Proposiciones 2.1 y 2.2).

(III) $\mathbb{R}^\infty = \{(a_n)_{n=1}^\infty : \text{sucesiones reales}\}$ es un e.v. con las operaciones

$$(a_n)_{n=1}^\infty + (b_n)_{n=1}^\infty = (a_n + b_n)_{n=1}^\infty$$
$$\alpha \cdot (a_n)_{n=1}^\infty = (\alpha a_n)_{n=1}^\infty$$

(IV) $\mathbb{P}^n_{\mathbb{R}}[x] = \{a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{R}\}$, es decir el conjunto de polinomios con coeficientes reales de grado menor o igual que $n$, en la variable $x$, es un e.v. con la suma y el producto habituales.

(V) $\mathcal{C}([a, b]) = \{f : [a, b] \longrightarrow \mathbb{R} : f \text{ continua}\}$ también es e.v. con la suma y producto habituales entre funciones.

(VI) $\mathbb{C}$ es un $\mathbb{C}$-espacio vectorial, y también es un $\mathbb{R}$-espacio vectorial, pero la estructura de ambos espacios es distinta, como se verá posteriormente.

Nótese que además de los típicos espacios $\mathbb{R}^n$ hay una multitud de espacios vectoriales de naturaleza bien distinta formados por elementos de los más diverso (funciones, polinomios, matrices, sucesiones, etc.). Lo más destacado del hecho de que estos conjuntos sean espacios vectoriales es que las propiedades que vamos a tratar en este tema y los siguientes son aplicables a la estructura de

[3] Insistimos en el hecho de que, a partir de ahora, cualquier elemento de un e.v. es un vector.

estos espacios, y por tanto funcionarán en todos ellos con independencia del tipo de conjunto con el que tratemos. Es por ello la necesidad de abordar de manera abstracta estos conceptos.

En contraste, veamos algunos conjuntos que no son espacios vectoriales.

**Ejemplo 4.2**

(I) $\mathbb{P}^n_{\mathbb{R}}[x] = \{a_0 + a_1x + \cdots + a_nx^n : a_i \in \mathbb{R}, a_n \neq 0\}$, es decir, el conjunto de polinomios con coeficientes reales de grado *exactamente* $n$ en la variable $x$, no es un e.v. puesto que el producto del escalar 0 por cualquier polinomio de grado exactamente $n$ nos proporciona el polinomio nulo, que ya no tiene grado exactamente $n$. Es decir, el producto por escalar con un elemento del conjunto no proporciona elementos del mismo conjunto, lo que lleva a que la dicha operación no esté correctamente definida en este conjunto.

(II) $\mathcal{M}^*_2(\mathbb{R}) = \{A \in \mathcal{M}_2(\mathbb{R}) : \det(A) = 0\}$, no es un espacio vectorial puesto que las matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ y } \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

pertenecen a dicho conjunto, pero no así su suma.

(III) $\mathbb{R}^\infty_l = \{(a_n)^\infty_{n=1}$ : sucesiones reales con límite $l\}$. Se propone al lector encontrar la causa. ¿Existe algún valor de $l$ para el cual este conjunto sí es un e.v.?

Quizás estos ejemplos lleven a confusión en el lector, pues todos ellos son subconjuntos de conjuntos mayores que sí son espacios vectoriales. Pero ahí está la clave de la comprensión de una estructura matemática: ésta es una propiedad del conjunto, no de algunos de sus elementos.

A continuación probaremos algunas propiedades simples que satisfacen los e.v.

**Proposición 4.1**

Sea $V$ un e.v. Se verifican las siguientes propiedades:

(I) El elemento neutro de un e.v. es único.

(II) El elemento opuesto de un e.v. es único.

(iii) $0 \cdot \mathbf{u} = \mathbf{0}$, $\forall \mathbf{u} \in V$.

(iv) El elemento opuesto de $\mathbf{u}$ es $(-1) \cdot \mathbf{u}$.

(v) $\alpha \cdot \mathbf{0} = \mathbf{0}$, $\forall \alpha \in \mathbb{K}$.

*Demostración:*
Si bien la demostración de estas propiedades es muy sencilla, el lector puede encontrarse con algunas dificultades en su comprensión, principalmente por no estar familiarizado con la abstracción matemática precisa para llevarlas a cabo. En cualquier caso, es cuestión de práctica conseguir entender los procedimientos habituales de demostración.

(i) Supongamos que $\mathbf{0}_1$ y $\mathbf{0}_2$ son dos elementos neutros y probemos que ambos elementos son en realidad el mismo. Debido a la propiedad de neutralidad:

$$\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_1 \text{ y } \mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2$$

Por la conmutatividad
$$\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2 + \mathbf{0}_1$$

y por tanto $\mathbf{0}_1 = \mathbf{0}_2$

(ii) Usamos la misma técnica que en el apartado anterior: si $\mathbf{v}_1$ y $\mathbf{v}_2$ son dos elementos opuestos de $\mathbf{u}$, entonces

$$(\mathbf{u} + \mathbf{v}_1) + \mathbf{v}_2 = (\mathbf{v}_2 + \mathbf{u}) + \mathbf{v}_1 = \mathbf{0} + \mathbf{v}_1 = \mathbf{v}_1$$
$$(\mathbf{u} + \mathbf{v}_2) + \mathbf{v}_1 = (\mathbf{u} + \mathbf{v}_1) + \mathbf{v}_2 = \mathbf{0} + \mathbf{v}_2 = \mathbf{v}_2$$

y como $(\mathbf{u} + \mathbf{v}_1) + \mathbf{v}_2 = (\mathbf{u} + \mathbf{v}_2) + \mathbf{v}_1$, se tiene que $\mathbf{v}_1 = \mathbf{v}_2$.

(iii) $\mathbf{u} = 1 \cdot \mathbf{u} = (0+1) \cdot \mathbf{u} = 0 \cdot \mathbf{u} + 1 \cdot \mathbf{u} = 0 \cdot \mathbf{u} + \mathbf{u}$. Nótese que toda esta cadena de igualdades se deduce de las propiedades que definen la estructura de e.v.

Como consecuencia de lo anterior $\mathbf{u} = \mathbf{u} + 0 \cdot \mathbf{u}$, luego $0 \cdot \mathbf{u}$ es el elemento neutro, es decir, $\mathbf{0}$.

(iv) $\mathbf{u} + (-1) \cdot \mathbf{u} = (1 + (-1)) \cdot \mathbf{u} = 0 \cdot \mathbf{u} = \mathbf{0}$, de donde se deduce que $(-1) \cdot \mathbf{u}$ es el opuesto de $\mathbf{u}$.

(v) $\alpha \cdot \mathbf{0} = \alpha \cdot (\mathbf{0} + \mathbf{0}) = \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0}$, luego $\alpha \cdot \mathbf{0} = \mathbf{0}$, y esto es válido $\forall \alpha \in \mathbb{K}$.

■

Nótese que el elemento nulo del e.v. es denotado genéricamente como $\mathbf{0}$, pero éste depende del tipo de espacio en el que nos encontremos. Por ejemplo, en $\mathbb{R}^n$, este elemento es el $(0,\dots,0)$, en $\mathcal{M}_{m\times n}(\mathbb{K})$ será la matriz nula y en $\mathbb{P}_R[x]$ será el polinomio nulo, esto es $p(x)=0$. En este último caso es importante que el lector entienda lo que esto significa: el polinomio nulo es aquel cuyos coeficientes son todos nulos, es decir, la expresión $p(x)=0$ no es una ecuación en la que $x$ es una incógnita, sino una igualdad que tiene que ser cierta para todo $x$, de modo que el polinomio que la verifica es el polinomio nulo.

## 4 2 INDEPENDENCIA LINEAL

**Definición 4.2**

Sea $V$ un e.v. y sean $\mathbf{v},\mathbf{v}_1,\dots,\mathbf{v}_n$ vectores de $V$. Se dice que $\mathbf{v}$ es *combinación lineal* (o que *depende linealmente*) de $\mathbf{v}_1,\dots,\mathbf{v}_n$, si existen $\alpha_1,\dots,\alpha_n\in\mathbb{K}$ tales que

$$\mathbf{v}=\alpha_1\mathbf{v}_1+\cdots+\alpha_n\mathbf{v}_n$$

Se dirá combinación lineal *no nula* si algún $\alpha_i\neq 0$.

Sin duda alguna, una de las principales ventajas del Álgebra Lineal es su simplicidad, debido precisamente a que las operaciones que podemos llevar a cabo con los vectores se reducen esencialmente a combinaciones lineales entre ellos. Como veremos, este concepto nos acompañará durante el resto de temas.[4]

**Proposición 4.2**

(I) El vector nulo es combinación lineal de cualquier conjunto de vectores.

(II) Un vector cualquiera $\mathbf{v}$ siempre es combinación lineal de cualquier conjunto de vectores que contenga al propio $\mathbf{v}$.

La demostración es muy sencilla: se trata de escribir el vector nulo en el primer caso, o el vector $\mathbf{v}$ en el segundo, como combinación lineal de cualesquiera

[4] Prácticamente todos los conceptos que se estudian en este tema aparecen, aunque de forma implícita, en la obra central de Grassmann *Die Lineale Ausdehnungslehre, ein neuer Zweig der Mathematik* de 1844. No es hasta 1910 cuando el matemático alemán Ernst Steinitz ofrece una presentación rigurosa y prácticamente definitiva de ellos.

otros vectores. ¿Se le ocurre al lector cómo hacerlo?

**Definición 4.3**

Se dice que los vectores $\mathbf{v}_1, \dots, \mathbf{v}_n$ son *linealmente dependientes (l.d.)* si podemos escribir el vector $\mathbf{0}$ como combinación lineal no nula de ellos. Dicho de otro modo, si existen escalares $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ *no todos nulos* tales que

$$\mathbf{0} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$$

En caso contrario se dirá que los vectores son *linealmente independientes (l.i.)*, lo que ocurrirá si cualquier combinación lineal de los vectores $\mathbf{v}_i$ igualada al vector nulo, implica que todos los escalares deben ser nulos, es decir,

$$\mathbf{0} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n \Longrightarrow \alpha_1 = \cdots = \alpha_n = 0$$

Probablemente el lector estará acostumbrado a entender la independencia lineal en un sentido geométrico (por ejemplo en $\mathbb{R}^2$, dos vectores son independientes si no son paralelos). Sin embargo el concepto se hace más difícil de entender si trabajamos con otro tipo de espacios. La siguiente proposición nos permite entender la definición anterior a través de una serie de propiedades de las que cabe destacar (i), la cual expresa a la perfección el significado de la dependencia lineal de vectores.

**Proposición 4.3**

(i) Si $\mathbf{v}_1, \dots, \mathbf{v}_n$ son vectores linealmente dependientes, existe algún $\mathbf{v}_j$ que es combinación lineal de los demás.

(ii) Todo conjunto finito de vectores entre los cuales se encuentre el vector $\mathbf{0}$ es linealmente dependiente.

(iii) Todo conjunto finito de vectores linealmente independientes no puede contener un subconjunto propio linealmente dependiente.

*Demostración:*

(i) Por la definición de dependencia lineal existe una combinación lineal no nula de estos vectores tal que

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}$$

Dado que la combinación lineal es no nula, debe existir al menos un $\alpha_j \neq 0$. Despejando de la expresión anterior $\mathbf{v}_j$ es claro que

$$\mathbf{v}_j = \frac{1}{\alpha_j}\left(-\alpha_1\mathbf{v}_1 - \cdots - \alpha_{j-1}\mathbf{v}_{j-1} - \alpha_{j+1}\mathbf{v}_{j+1} - \cdots - \alpha_n\mathbf{v}_n\right)$$

de modo que $\mathbf{v}_j$ es combinación lineal del resto.

(II) Obviamente, si el vector nulo forma parte de un conjunto de vectores, existe una combinación lineal no nula de estos (la formada por todos los escalares nulos, menos el que multiplica al vector cero) que proporciona el vector nulo. Luego el conjunto es l.d.

(III) Para probar este hecho usaremos una técnica de demostración conocida como *reducción al absurdo*, la cual consiste en suponer lo contrario de lo que queremos demostrar y deducir de aquí una contradicción, lo que prueba que nuestra suposición es falsa, y por tanto el resultado es cierto.

Supongamos entonces que tenemos un conjunto de vectores $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ que es l.i. en el que existe un subconjunto propio que es l.d. Por simplicidad en la notación pongamos que el conjunto l.d. está formado por los primeros $k$ vectores, con $k < n$. Puesto que estos vectores son l.d., existe una combinación lineal no nula,

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

con algún $\alpha_j \neq 0$. Entonces es evidente que

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k + 0 \cdot \mathbf{v}_{k+1} + \cdots + 0 \cdot \mathbf{v}_n = \mathbf{0}$$

Esto es una combinación lineal de los vectores $\mathbf{v}_1, \dots, \mathbf{v}_n$ igualada al vector nulo, pero en el que hay escalares no nulos, lo cual es imposible, pues estos vectores son l.i. De esta contradicción se sigue que nuestra suposición sobre la existencia de un subconjunto propio l.d. es falsa.

■

El siguiente resultado nos proporciona un método sencillo para detectar la dependencia o independencia de vectores de $\mathbb{K}^n$ (y como luego veremos, también se podrá usar en otros espacios).

**Teorema 4.1**

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ con rango$(A) = r$. Entonces existen $r$ filas (o columnas) de $A$ linealmente independientes, esto es, hay $r$ vectores de $\mathbb{K}^n$ correspondientes a sus filas (o de $\mathbb{K}^m$ correspondientes a sus columnas) linealmente independientes, de manera que el resto se expresa como combinación lineal de éstas.

*Demostración:*
Consideremos $A = (a_{ij})$ una matriz de rango $r$. Para simplificar la notación supongamos que un menor de orden $r$ no nulo se obtiene con las primeras $r$ filas y columnas,[5] es decir,

$$\begin{vmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} \end{vmatrix} \neq 0 \tag{4.1}$$

Dividiremos la demostración en dos pasos.

(I) Probemos en primer lugar que los vectores de $\mathbb{K}^n$ correspondientes a las $r$ primeras filas de $A$, es decir,

$$\mathbf{a}_1 = (a_{11}, \dots, a_{1n}), \ \dots, \ \mathbf{a}_r = (a_{r1}, \dots, a_{rn})$$

son l.i.

Procederemos por reducción al absurdo. Supongamos que dicho conjunto de vectores es l.d. Por (I) de la Proposición 4.3 debe existir uno de ellos que sea combinación lineal (no nula) de los restantes. Supongamos que es el $\mathbf{a}_j$. Entonces,

$$\mathbf{a}_j = \alpha_1 \mathbf{a}_1 + \cdots + \alpha_{j-1}\mathbf{a}_{j-1} + \alpha_{j+1}\mathbf{a}_{j+1} + \cdots + \alpha_r \mathbf{a}_r$$

Realicemos ahora la siguiente operación en el menor básico dado en (4.1)

$$F_j - \alpha_1 F_1 - \cdots - \alpha_{j-1}F_{j-1} - \alpha_{j+1}F_{j+1} - \cdots - \alpha_r F_r$$

Puesto que las filas del menor coinciden con los vectores $\mathbf{a}_i$, obtenemos que la fila $j$-ésima del menor es nula, y por tanto su determinante es cero, lo que es una contradicción con nuestra hipótesis inicial, y así los vectores deben ser l.i.

(II) Veamos que el resto de filas depende linealmente de las $r$ primeras. Consideremos $k$ y $l$ tales que $r < k \leq m$, $1 \leq l \leq n$, y el menor básico de (4.1) al que orlamos con la fila $k$-ésima y la columna $l$-ésima, esto es

$$\begin{vmatrix} a_{11} & \cdots & a_{1r} & a_{1l} \\ \vdots & \ddots & \vdots & \vdots \\ a_{r1} & \cdots & a_{rr} & a_{rl} \\ a_{k1} & \cdots & a_{kr} & a_{kl} \end{vmatrix}$$

Esta claro que si $l \leq r$ el determinante es nulo pues tendría dos columnas iguales; y si $l > r$, entonces correspondería a un menor de orden $r+1$, que también es nulo gracias a la hipótesis inicial (esto es, rango$(A) = r$).

[5] Esto no supone restricción alguna puesto que sabemos que el rango de una matriz no se altera por intercambio de filas y/o columnas.

Desarrollemos ahora este determinante por su última columna. Tendremos

$$0 = (-1)^{r+2} a_{1l}|A_1| + \cdots + a_{kl}|A_{r+1}|$$

Ahora bien, observemos que los $A_i$ son siempre iguales, con independencia del $l$ escogido, y además $|A_{r+1}| \neq 0$. Por tanto, para cada $l = 1, \ldots, n$, podemos escribir (despejando $a_{kl}$)

$$a_{kl} = (-1)^{r+1} a_{1l} \frac{|A_1|}{|A_{r+1}|} + \cdots + a_{rl} \frac{|A_r|}{|A_{r+1}|}$$

pero esto es lo mismo que

$$\left.\begin{array}{rcl} a_{k1} & = & \alpha_1 a_{11} + \cdots + \alpha_r a_{r1} \\ & \vdots & \\ a_{kn} & = & \alpha_1 a_{1n} + \cdots + \alpha_r a_{rn} \end{array}\right\}$$

donde $\alpha_i = (-1)^{r+i} \frac{|A_i|}{|A_{r+1}|}$. De aquí se deduce que $\mathbf{a}_k$ es combinación lineal de $\mathbf{a}_1, \ldots, \mathbf{a}_r$, y esto es cierto para cada $k$, $r < k \leq n$.

Un argumento similar se emplea para la demostración por columnas.

■

**Nota 4.1**

Como consecuencia de este resultado podemos probar ahora la suficiencia en el Teorema de Rouché-Frobenius (Teorema 3.2).

Si rango$(\bar{A})$ = rango$(A)$, eso significa, en virtud del Teorema 4.1, que la última columna de la matriz $\bar{A}$, es decir el vector $\mathbf{b}$, es combinación lineal de las $n$ primeras columnas de $A$

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Esto es, existen escalares $x_1, \ldots, x_n \in \mathbb{K}$ tales que

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

es decir,

$$\left.\begin{array}{rcl} b_1 & = & a_{11}x_1 + \cdots + a_{1n}x_n \\ & \vdots & \\ b_m & = & a_{m1}x_1 + \cdots + a_{mn}x_n \end{array}\right\}$$

luego el sistema es compatible.

**Ejemplo 4.3**

(I) Estudiar la dependencia o independencia lineal de los vectores de $\mathbb{R}^3$: $\mathbf{v}_1 = (1,1,3)$, $\mathbf{v}_2 = (0,1,2)$ y $\mathbf{v}_3 = (1,2,5)$.

Como hemos visto, basta estudiar el rango de la matriz cuyas filas (o columnas) corresponden a las coordenadas de los vectores dados, i.e., nos preguntamos por el rango de la matriz

$$\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 1 & 2 & 5 \end{pmatrix}$$

Como

$$\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \neq 0 \quad \text{y} \quad \begin{vmatrix} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 1 & 2 & 5 \end{vmatrix} = 0$$

los vectores $\mathbf{v}_1$ y $\mathbf{v}_2$ son linealmente independientes y el vector $\mathbf{v}_3$ es combinación lineal de ellos.

Nótese que también se tiene que los vectores $\mathbf{v}_2$ y $\mathbf{v}_3$ son independientes (y $\mathbf{v}_1$ es combinación de ellos), pues la matriz formada por las dos últimas filas también tiene rango dos.

(II) Comprobar que el conjunto de vectores de $\mathbb{P}^n_{\mathbb{R}}[x]$, $\{1, x, \dots, x^n\}$, es linealmente independiente.

Nótese que en este caso no podemos construir matriz alguna, por lo que es preciso acudir a la definición de independencia lineal. Es decir, si consideramos una combinación lineal de estos vectores igualada al vector nulo, debemos deducir que los escalares son todos nulos. En efecto,

$$\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n = 0$$

es una igualdad entre polinomios (¡no una ecuación en $x$!, pues el vector nulo de los polinomios es el polinomio cero). La única posibilidad de que estos dos polinomios sean iguales es que sus coeficientes lo sean, luego $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$, es decir los escalares son todos nulos, lo cual prueba la independencia.

(III) Comprobar que el conjunto de vectores de $\mathcal{C}([0,2\pi])$, $\{1, \operatorname{sen}^2 x, \cos^2 x\}$, es l.d.

Sabemos por la fórmula fundamental de la trigonometría que $\cos^2 x + \operatorname{sen}^2 x - 1 = 0$. Luego tenemos una combinación lineal no nula de estos vectores igualada al vector cero de este espacio (la función nula).

**Definición 4.4**

Se denomina *rango de un conjunto de vectores* al mayor número de ellos que son linealmente independientes.

Lógicamente, en el caso de vectores de $\mathbb{K}^n$, el rango de un conjunto de vectores coincide con el de la matriz que forman.

Una primera consecuencia del Teorema 4.1 es la siguiente:

**Teorema 4.2**

En $\mathbb{K}^n$, todo conjunto de vectores formado por $n+1$ vectores es linealmente dependiente.

La demostración es inmediata, pues el rango de una matriz formada por $n+1$ vectores de $\mathbb{K}^n$ es, como máximo, $n$. Al no poder tener rango $n+1$, los vectores no pueden ser independientes.

## 4.3 BASES Y DIMENSIÓN DE UN ESPACIO VECTORIAL

El concepto de base es posiblemente el elemento más importante con el que nos encontramos en un espacio vectorial. Antes de llegar a él precisamos de la siguiente definición.

**Definición 4.5**

Sea $V$ un e.v. Un conjunto de vectores $\mathbf{v}_1, \dots, \mathbf{v}_n$ se dice *sistema generador* (o *conjunto generador*) de $V$ si cualquier vector $\mathbf{u} \in V$ se puede poner como combinación lineal de ellos.

En cierto modo, podemos entender un sistema generador como la "semilla" o los "padres" de un espacio vectorial, de tal modo que, realizando combinaciones lineales con sus elementos somos capaces de construir cualquier otro vector del espacio.

**Ejemplo 4.4**

(i) En $\mathbb{R}^2$, el conjunto $\{(1,0),(0,1)\}$ es un sistema generador. Para comprobarlo escogemos un vector arbitrario de $\mathbb{R}^2$, $(x_1, x_2)$ y tratamos de escribirlo como combinación lineal de los anteriores. Esto es,

$$(x_1, x_2) = \alpha(1,0) + \beta(0,1)$$

Un simple cálculo nos demuestra que $\alpha = x_1$ y $\beta = x_2$. Es decir, dado cualquier vector somos capaces de encontrar una combinación lineal de elementos de nuestro conjunto que nos permiten obtener el vector dado.

(ii) El conjunto formado por $(1,0)$ y $(2,0)$ no es sistema generador de $\mathbb{R}^2$, pues si tratamos de repetir lo anterior

$$(x_1, x_2) = \alpha(1,0) + \beta(2,0),$$

se obtiene que $x_2 = 0$. Es decir, no cualquier vector de $\mathbb{R}^2$ es combinación lineal de los anteriores (lo son únicamente aquellos vectores cuya segunda componente es cero), y por tanto no generan todo el espacio.

El concepto de sistema generador tiene una enorme trascendencia, pues nos va a permitir estudiar las propiedades de un espacio "mirando" solo a sus generadores. Sin embargo adolece de un pequeño problema que se muestra en el siguiente resultado.

**Lema 4.4**

Sea $\mathcal{S}$ un sistema generador de un e.v. $V$. Si $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$, con $\mathcal{S}_1$ y $\mathcal{S}_2$ conjuntos disjuntos[6] tales que los elementos de $\mathcal{S}_2$ se escriben como combinación lineal de los elementos de $\mathcal{S}_1$, entonces $\mathcal{S}_1$ es sistema generador.

*Demostración:*
Consideremos $\mathcal{S} = \{\mathbf{u}_1, \dots, \mathbf{u}_l, \mathbf{u}_{l+1}, \dots, \mathbf{u}_m\}$, donde

$$\mathcal{S}_1 = \{\mathbf{u}_1, \dots, \mathbf{u}_l\}, \quad \mathcal{S}_2 = \{\mathbf{u}_{l+1}, \dots, \mathbf{u}_m\}$$

y supongamos que cada $\mathbf{u}_j$, $l+1 \leq j \leq m$ es combinación lineal de los vectores de $\mathcal{S}_1$, es decir,

$$\mathbf{u}_j = \sum_{i=1}^{l} x_{ij}\mathbf{u}_i$$

Como $\mathcal{S}$ es sistema generador, dado $\mathbf{u} \in V$ cualquiera, $\mathbf{u}$ es combinación lineal de los $\mathbf{u}_i$, $1 \leq i \leq m$. Entonces,

$$\begin{aligned}
\mathbf{u} &= \sum_{j=1}^{m} \alpha_j \mathbf{u}_j = \sum_{j=1}^{l} \alpha_j \mathbf{u}_j + \sum_{j=l+1}^{m} \alpha_j \mathbf{u}_j = \sum_{j=1}^{l} \alpha_j \mathbf{u}_j + \sum_{j=l+1}^{m} \alpha_j \left( \sum_{i=1}^{l} x_{ij}\mathbf{u}_i \right) \\
&= \sum_{j=1}^{l} \alpha_j \mathbf{u}_j + \sum_{i=1}^{l} \left( \sum_{j=l+1}^{m} \alpha_j x_{ij} \right) \mathbf{u}_i = \sum_{k=1}^{l} \left( \alpha_k + \sum_{j=l+1}^{m} \alpha_k x_{kj} \right) \mathbf{u}_k
\end{aligned}$$

es decir, $\mathbf{u}$ es combinación lineal de los elementos de $\mathcal{S}_1$. Como el razonamiento se ha hecho para un vector $\mathbf{u}$ arbitrario, significa que es cierto para todos, y por tanto $\mathcal{S}_1$ es sistema generador.

■

Lo que prueba este resultado es que si tenemos un conjunto que es sistema generador en el cual existe uno o más vectores que son combinación lineal de otros vectores del conjunto, podemos suprimirlos y seguimos teniendo un sistema generador. Dicho de otro modo, en los sistemas generadores podemos encontrarnos vectores que no son necesarios para generar el espacio, pues si los eliminamos, seguimos teniendo un sistema generador. Esto plantea un inconveniente a la hora de trabajar con un sistema generador, pues puede haber vectores en el mismo que sean innecesarios. Para resolver este pequeño problema aparece el concepto de base.

[6] Esto es, $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$.

**Definición 4.6**

Sea $V$ un e.v. Un conjunto finito de vectores $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ es una *base* de $V$ si es un conjunto linealmente independiente y sistema generador.

A la vista de lo comentado anteriormente, para evitar los problemas de "exceso de vectores" en los sistemas generadores lo que hacemos simplemente es quedarnos con aquellos que sean linealmente independientes (recuérdese (III) de la Proposición 4.3).

**Ejemplo 4.5**

(I) En $\mathbb{R}^n$, el conjunto $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ donde $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ (un 1 en la componente $i$) es una base de $\mathbb{R}^n$.

En efecto, para ver que es l.i. sólo debemos estudiar el rango de la matriz formada por los vectores. Es inmediato comprobar que dicha matriz es la identidad, que tiene rango $n$, por lo que el conjunto es l.i.

Para ver que es sistema generador tratamos de escribir un vector arbitrario de $\mathbb{R}^n$ como combinación lineal de ellos:

$$(x_1, \ldots, x_n) = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n$$

Un simple cálculo nos lleva a que $x_i = \alpha_i$, $\forall i$. Luego este conjunto también es sistema generador.

(II) El conjunto $\{1, x, \ldots, x^n\}$ es una base de $\mathbb{P}^n_{\mathbb{R}}[x]$.

(III) El conjunto

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

es una base de $\mathcal{M}_2(\mathbb{R})$.

Se propone al lector que demuestre que los dos últimos conjuntos son base de sus respectivos espacios.

**Nota 4.2**

Todas las bases que aparecen en el ejemplo 4.5 son denominadas *bases*

*canónicas*, por la especial simplicidad con la que generan el espacio. Obsérvese la facilidad en la comprobación de que cada uno de ellos es sistema generador.

**Teorema 4.3**

Sea $\mathcal{B} = \{\mathbf{u}_1, \dots \mathbf{u_n}\}$ una base de un e.v. $V$. Entonces $\forall \mathbf{u} \in V$ existen unos únicos $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ tales que

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n \tag{4.2}$$

*Demostración:*
Puesto que $\mathcal{B}$ es una base, en particular es un sistema generador, y por tanto la existencia de los $\alpha_i$ está trivialmente garantizada.

Veamos que son únicos. Para ello supongamos que existen $\alpha_i$, $\beta_i$, $1 \leq i \leq n$ tales que

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n$$

$$\mathbf{u} = \beta_1 \mathbf{u}_1 + \cdots + \beta_n \mathbf{u}_n$$

Restando ambas expresiones,

$$\mathbf{0} = (\alpha_1 - \beta_1)\mathbf{u}_1 + \cdots + (\alpha_n - \beta_n)\mathbf{u}_n$$

Esto es una combinación lineal de los elementos de $\mathcal{B}$ igualada al vector cero. Como el conjunto $\mathcal{B}$ es l.i., se tiene que

$$\alpha_1 - \beta_1 = \cdots = \alpha_n - \beta_n = 0$$

Luego $\alpha_i = \beta_i \ \forall i$, de donde se tiene la unicidad.

■

**Definición 4.7**

A los escalares $\alpha_1, \dots, \alpha_n$ de (4.2) se les denominan *coordenadas* de $\mathbf{u}$ en la base $\mathcal{B}$, y se notará por

$$\mathbf{u}_{\mathcal{B}} = (\alpha_1, \dots, \alpha_n)_{\mathcal{B}}$$

Cuando no haya posibilidad de confusión omitiremos el subíndice.

La trascendencia de este resultado es enorme, pues lo que nos asegura es que las bases son sistemas generadores que generan el espacio de forma única. Es decir, cada vector tiene asociado un único "identificador" (una especie de D.N.I.), que son sus coordenadas, que corresponden a los escalares de la combinación lineal respecto de la cual escribimos el vector.

**Ejemplo 4.6**

(i) Consideremos el vector $(1,-2,3) \in \mathbb{R}^3$. ¿Cuáles son sus coordenadas respecto de la base canónica de $\mathbb{R}^3$, $\mathcal{B}_c = \{(1,0,0),(0,1,0),(0,0,1)\}$? Bastará observar que

$$(1,-2,3) = 1 \cdot (1,0,0) + (-2) \cdot (0,1,0) + 3 \cdot (0,0,1)$$

En general, las coordenadas de un vector $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n$ respecto de la base canónica de este espacio son precisamente $(x_1,\dots,x_n)$.

(ii) ¿Cuáles son las coordenadas del vector $(1,-1,0)$ respecto de la base $\mathcal{B} = \{(2,0,1),(1,1,2),(0,0,-1)\}$? En este caso hemos de encontrar unos escalares $\alpha_1, \alpha_2, \alpha_3$ tales que

$$(1,-1,0) = \alpha_1(2,0,1)+\alpha_2(1,1,2)+\alpha_3(0,0,-1) \Rightarrow \left\{ \begin{array}{l} 2\alpha_1 + \alpha_2 = 1 \\ \alpha_2 = -1 \\ \alpha_1 + 2\alpha_2 - \alpha_3 = 0 \end{array} \right.$$

de donde se obtiene $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = -1$, es decir $(1,-1,0) = (1,-1,-1)_{\mathcal{B}}$

(iii) ¿Cuáles son las coordenadas del polinomio $1 + 3x - 2x^2 \in \mathbb{P}^2_{\mathbb{R}}[x]$ respecto de la base canónica $\{1, x, x^2\}$. Nuevamente escribimos

$$1 + 3x - 2x^2 = 1 \cdot 1 + 3 \cdot x + (-2) \cdot x^2$$

resultando que las coordenadas son $(1,3,-2)$. Nótese cómo los vectores de $\mathbb{P}^2_{\mathbb{R}}[x]$ pueden identificarse con vectores de $\mathbb{R}^3$. Recíprocamente, si nos dan un vector de $\mathbb{P}^2_{\mathbb{R}}[x]$ cuyas coordenadas respecto de la base canónica son $(-1,0,2)$, en realidad nos están dando el polinomio $2x^2-1$. Es fundamental resaltar la importancia del orden de los vectores de una base. Si en lugar de usar la base canónica $\{1, x, x^2\}$ estamos usando la base $\{x^2, x, 1\}$ (¡que no es la misma!), entonces el vector $(-1,0,2)$ correspondería al polinomio $2 - x^2$.

**Nota 4.3**

Gracias al uso de coordenadas, estudiar la dependencia lineal de un conjunto de vectores que no estén en $\mathbb{K}^n$ resulta más sencillo que lo que hicimos en (II) del ejemplo 4.3. Por ejemplo, para ver si el conjunto $\{x^2+1, 1+x^3, -x\} \subset \mathbb{P}^3_{\mathbb{R}}[x]$ es o no l.i., podemos simplemente escribir los vectores mediante sus coordenadas respecto de la base canónica, resultando $\{(1,0,1,0),(1,0,0,1),(0,-1,0,0)\}$ y estudiar el correspondiente rango de la matriz (como si fueran vectores en $\mathbb{R}^4$):

$$\text{rango}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \end{pmatrix} = 3$$

lo que indica que son independientes.

Veamos a continuación algunos consecuencias interesantes del hecho de que existan bases en un e.v.

**Proposición 4.5**

Supongamos que el e.v. $V$ posee una base formada por $n$ elementos. Entonces, todo conjunto de $m$ vectores, con $m > n$ es l.d.

*Demostración:*

Sea $\{\mathbf{u}_1, \dots \mathbf{u_n}\}$ una base de $V$. Sean $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}$, $n+1$ vectores de $V$. Por el teorema anterior, cada $\mathbf{x}_i$ es combinación lineal única de la base, luego

$$\mathbf{x}_i = \sum_{j=1}^{n} x_{ij}\mathbf{u}_j, \quad i = 1, \dots, n+1 \tag{4.3}$$

Veamos que estos $n+1$ vectores son l.d. Para ello consideramos una combinación lineal igualada al vector nulo:

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n + \alpha_{n+1}\mathbf{x}_{n+1} = \mathbf{0}$$

Usando (4.3),

$$\alpha_1\left(\sum_{j=1}^{n} x_{1j}\mathbf{u}_j\right) + \cdots + \alpha_n\left(\sum_{j=1}^{n} x_{nj}\mathbf{u}_j\right) + \alpha_{n+1}\left(\sum_{j=1}^{n} x_{n+1,j}\mathbf{u}_j\right) = \mathbf{0}$$

Agrupando términos,

$$\left(\sum_{j=1}^{n+1} \alpha_j x_{j1}\right)\mathbf{u}_1 + \cdots + \left(\sum_{j=1}^{n+1} \alpha_j x_{jn}\right)\mathbf{u}_n = \mathbf{0}$$

Como $\mathbf{u}_1, \dots, \mathbf{u}_n$ forman base, son l.i., por tanto cada uno de los escalares de la anterior combinación lineal es nulo, esto es

$$\left.\begin{array}{c} \alpha_1 x_{11} + \cdots + \alpha_n x_{n1} + \alpha_{n+1} x_{n+1,1} = 0 \\ \vdots \\ \alpha_1 x_{1n} + \cdots + \alpha_n x_{nn} + \alpha_{n+1} x_{n+1,n} = 0 \end{array}\right\}$$

Ahora bien, esto es un sistema homogéneo de $n$ ecuaciones con $n+1$ incógnitas (los $\alpha_i$). Como el rango de la matriz de los coeficientes es estrictamente menor que el número de incógnitas, es un sistema compatible indeterminado; por tanto posee solución distinta de la trivial. Es decir, existen $\alpha_i$ no todos nulos tales que se verifica (4.3). Luego el conjunto formado por $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}$ es l.d.

Así, hemos probado que cualquier conjunto de $n+1$ vectores siempre es l.d., luego si $m > n$, cualquier conjunto de $m$ vectores también es l.d. (cf. (iii) de la Proposición 4.3).

■

Este resultado admite una prueba más sencilla si usamos coordenadas respecto de una base y el Teorema 4.1. ¿Podría el lector llevarla a cabo?

Como consecuencia inmediata de este resultado, se tiene el siguiente:

**Teorema 4.4**

Todas las bases de un e.v. poseen el mismo número de elementos.

**Definición 4.8**

Se llama *dimensión* de un espacio vectorial $V$ al número de elementos de cualquiera de sus bases, y se notará por $\dim(V)$.

**Nota 4.4**

Por convenio se considera que el espacio vectorial $V = \{\mathbf{0}\}$ tiene dimensión cero.

**Ejemplo 4.7**

Atendiendo al ejemplo 4.5, es inmediato obtener las dimensiones de los espacios más comunes:

(I) $\dim(\mathbb{P}_{\mathbb{R}}^n[x]) = n + 1$.

(II) $\dim(\mathbb{K}^n) = n$.

(III) $\dim(\mathcal{M}_{m\times n}(\mathbb{K})) = m \cdot n$.

Una vez que hemos establecido la importancia de las bases veamos cómo de una forma sencilla podemos comprobar que un conjunto es base de un e.v.

**Proposición 4.6**

Sea $V$ es un e.v. de dimensión $n$. Todo conjunto de $n$ vectores l.i. forma una base de $V$.

*Demostración:*
Puesto que $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ es un conjunto l.i. de $V$, para probar que es una base sólo es necesario comprobar que se trata de un sistema generador.

Sea entonces $\mathbf{u} \in V$. Por la Proposición 4.5, el conjunto $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}\}$ es l.d. Por definición de dependencia lineal existen $\alpha_1, \dots, \alpha_n, \alpha_{n+1}$ no todos nulos tales que

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n + \alpha_{n+1}\mathbf{u} = \mathbf{0}$$

De todos ellos, $\alpha_{n+1}$ no puede ser nulo, pues si lo fuera, tendríamos una combinación de los $\mathbf{v}_i$ igualada al vector cero, y como éstos son l.i., todos los escalares tendrían que anularse. Al ser $\alpha_{n+1} \neq 0$, podemos escribir

$$\mathbf{u} = -\frac{\alpha_1}{\alpha_{n+1}}\mathbf{v}_1 - \cdots - \frac{\alpha_n}{\alpha_{n+1}}\mathbf{v}_n$$

Es decir, hemos escrito un vector arbitrario $\mathbf{u}$ como combinación lineal de los $\mathbf{v}_i$, por tanto son sistema generador, como queríamos demostrar.

■

La consecuencia de este resultado es bien simple: para comprobar que un conjunto es base sólo necesitamos que sea l.i. y que tenga tantos elementos como la dimensión del espacio en el que nos encontremos; automáticamente será sistema generador y por tanto base.

Para finalizar esta sección veremos una serie de resultados que nos dicen qué podemos hacer cuando tenemos conjuntos con más o menos vectores que la

dimensión del espacio.

**Proposición 4.7 (Ampliación de bases)**

Sea $V$ un espacio vectorial de $\dim V = n$. Si $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ con $k < n$, es un conjunto l.i. entonces existen $n-k$ vectores, $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$, tales que el conjunto $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ es una base de $V$.

*Demostración:*
Como $k < n$ debe existir al menos un vector $\mathbf{v}_{k+1}$ que sea l.i. con $\mathbf{v}_1, \dots, \mathbf{v}_k$ (en caso contrario cualquier vector se escribiría como combinación lineal de ellos, y por tanto tendríamos un conjunto l.i. que es s.g., es decir una base). Repetimos el proceso añadiendo un nuevo vector $\mathbf{v}_{k+2}$ al conjunto l.i. $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}\}$. El proceso termina cuando juntemos $n$ vectores, ya que tendríamos $n$ vectores l.i. en un espacio de dimensión $n$.

■

**Ejemplo 4.8**

Dado el conjunto de $\mathbb{R}^3$, $(1,0,1), (2,1,1)$, ampliarlo hasta conseguir una base de este espacio.

Comprobamos en primer lugar que son l.i. Para ello sólo es necesario estudiar el rango de la matriz correspondiente

$$\text{rango}\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix} = 2 \quad \text{pues} \quad \begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} \neq 0$$

Para ampliar a una base, precisamos $3-2=1$ vector que sea l.i. con los anteriores, es decir, tal que el rango de la matriz

$$\text{rango}\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} = 3$$

En este punto, es evidente que podemos tomar infinitos vectores distintos que completan a una base. En tales casos conviene tomar vectores “sencillos”, que tengan un gran número de ceros, Una posibilidad sería el vector $(0,0,1)$. Nótese que esta elección se ha hecho en función de un menor de orden dos distinto de cero formado por los vectores dados.

**Proposición 4.8**

Si $\mathcal{S}$ es un sistema generador de un e.v. $V$ de dimensión $n$, entonces existe $\mathcal{S}_1$ un subconjunto de $\mathcal{S}$ que es base de $V$.

*Demostración:*
Sea $\mathbf{u}_1 \in \mathcal{S}$, $\mathbf{u}_1 \neq \mathbf{0}$ (siempre existirá tal vector, a menos que $V = \{\mathbf{0}\}$, en cuyo caso no hay nada que probar). Consideremos ahora $\mathbf{u}_2 \in \mathcal{S}$ l.i. con $\mathbf{u}_1$ (un tal elemento siempre existe, pues en caso contrario el resto de elementos sería combinación lineal, y por tanto $\mathbf{u}_1$ ya sería sistema generador, y l.i., es decir, una base). Reiteramos el procedimiento hasta obtener un conjunto de vectores de $\mathcal{S}$ l.i. Por el comentario anterior, este conjunto será una base de $V$.

■

El siguiente corolario resume los resultados anteriores:

**Corolario 4.5**

Si $V$ es un e.v. con $\dim V = n$, se tiene:

(I) Todo conjunto de $n$ vectores l.i. es una base.

(II) Todo conjunto con más de $n$ vectores es l.d.

(III) Todo sistema generador tiene al menos $n$ elementos.

(IV) Todo sistema generador de $n$ elementos es una base.

Los resultados anteriores hacen referencia a las propiedades de las bases, pero en ningún momento hemos establecido que tales bases existan en cualquier espacio vectorial. De hecho, hemos de distinguir entre dos tipos de espacios, aquellos en los que existen bases, que son los espacios de *dimensión finita*, y espacios en los que no existen bases, que son los *espacios de dimensión infinita*, que definimos a continuación.

**Definición 4.9**

Un conjunto infinito de vectores de un e.v. $V$ es l.i. si cualquier subconjunto suyo es l.i. Si existe un tal conjunto se dirá que $V$ es de *dimensión infinita*.

**Ejemplo 4.9**

(I) $\mathbb{P}_{\mathbb{R}}[x]$ = {polinomios con coeficientes reales en la variable $x$} es un espacio vectorial de dimensión infinita. Basta comprobar que la familia

$$1, x, x^2, \dots, x^n, \dots$$

es linealmente independiente.

(II) En $\mathcal{C}([0, 2\pi])$, la familia $\{\operatorname{sen}(nx), \cos(nx)\}_{n\in\mathbb{N}}$ es l.i., luego $\mathcal{C}([0, 2\pi])$ es un espacio de dimensión infinita.

**Nota 4.5**

Es importante resaltar que en los espacios de dimensión infinita no existen coordenadas, pues no hay bases.

La importancia de los espacios de dimensión infinita es grande, sin embargo constituyen un paso más en el estudio de estructuras matemáticas, por lo que prácticamente no son considerados a lo largo del texto.

## 4 4 CAMBIOS DE BASE

Si atendemos a (II) del ejemplo 4.6 vemos que un mismo vector tiene coordenadas distintas respecto de diferentes bases. En esta sección pretendemos relacionar las coordenadas de un vector cuando usamos bases distintas. Para ello consideremos dos bases de un e.v. $V$ de dimensión $n$:

$$\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}, \qquad \mathcal{B}' = \{\mathbf{e}'_1, \dots, \mathbf{e}'_n\}$$

Cada uno de los vectores de $\mathcal{B}'$ se puede escribir como combinación lineal de los vectores de $\mathcal{B}$ de forma única (cf. Teorema 4.3), es decir,

$$\left.\begin{array}{c} \mathbf{e}'_1 = a_{11}\mathbf{e}_1 + \cdots + a_{n1}\mathbf{e}_n \\ \vdots \\ \mathbf{e}'_n = a_{1n}\mathbf{e}_1 + \cdots + a_{nn}\mathbf{e}_n \end{array}\right\} \quad \text{i.e. } \mathbf{e}'_j = \sum_{i=1}^{n} a_{ij}\mathbf{e}_i \tag{4.4}$$

Con las coordenadas de cada uno de los vectores de $\mathcal{B}'$ construimos la matriz $A \in \mathcal{M}_n(\mathbb{K})$, dada por $A = (a_{ij})$. Es importante observar cómo está construida

dicha matriz: **la columna $j$-ésima de $A$ corresponde a las coordenadas del vector $\mathbf{e}'_j$ respecto de la base $\mathcal{B}$**. A esta matriz se le denomina *matriz del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$*, y se notará por $M^{\mathcal{B}}_{\mathcal{B}'}$ ó $M(\mathcal{B}';\mathcal{B})$.

¿Qué relación mantienen las coordenadas de un vector respecto de ambas bases?
Escribamos las coordenadas de un vector respecto de la base $\mathcal{B}$ por $\mathbf{x}_{\mathcal{B}}$ y sus coordenadas respecto de $\mathcal{B}'$ por $\mathbf{x}_{\mathcal{B}'}$. Es decir,

$$\mathbf{x}_{\mathcal{B}} = \sum_{i=1}^{n} x_i \mathbf{e}_i, \quad \mathbf{x}_{\mathcal{B}'} = \sum_{j=1}^{n} x'_j \mathbf{e}'_j \tag{4.5}$$

Usando (4.4),

$$\mathbf{x}_{\mathcal{B}'} = \sum_{j=1}^{n} x'_j \left( \sum_{i=1}^{n} a_{ij} \mathbf{e}_i \right) = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} x'_j a_{ij} \right) \mathbf{e}_i$$

Comparando el último término de la expresión anterior con (4.5) (puesto que las coordenadas son únicas), resulta que

$$x_i = \sum_{j=1}^{n} a_{ij} x'_j$$

Si ahora recordamos la definición del producto de matrices, resulta que

$$\mathbf{x}_{\mathcal{B}} = M^{\mathcal{B}}_{\mathcal{B}'} \mathbf{x}_{\mathcal{B}'} \tag{4.6}$$

que son las denominadas *ecuaciones del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$*.

**Nota 4.6**

La notación de las ecuaciones del cambio de base puede ser un tanto confusa pues estamos habituados a escribir los vectores $\mathbf{x} = (x_1, \ldots, x_n)$ como una matriz fila. Sin embargo, la expresión (4.6) sólo tiene sentido cuando expresamos $\mathbf{x}_{\mathcal{B}}$ y $\mathbf{x}_{\mathcal{B}'}$ como vectores columna (para poder hacer correctamente la multiplicación matricial). En todo lo que sigue, en cualquier operación que involucre la multiplicación matricial de vectores consideraremos éstos como matrices columna, de modo que cuando tengan que ser multiplicados por filas usaremos su traspuesta, es decir, $\mathbf{x}^T$. Cuando no haya multiplicaciones, seguiremos escribiéndolos por filas.

**Teorema 4.6**

Con las notaciones anteriores, la matriz del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$ es invertible y su inversa es la matriz del cambio de base de $\mathcal{B}$ a $\mathcal{B}'$.

*Demostración:*
Puesto que los vectores de $\mathcal{B}'$ son l.i., ya que forman base, si escribimos la matriz de sus coordenadas respecto de la base $\mathcal{B}$, esto es la matriz $A$, debemos tener que $\operatorname{rango}(A) = n$. Esto es, $|A| \neq 0$, y por tanto $A$ es invertible.

Sea ahora $A' = M_{\mathcal{B}}^{\mathcal{B}'}$, con $A' = (a'_{ij})$, donde, por definición,

$$\mathbf{e}_j = \sum_{i=1}^{n} a'_{ij}\mathbf{e}'_i$$

Usando (4.4) se tiene que

$$\mathbf{e}_j = \sum_{i=1}^{n} a'_{ij}\left(\sum_{k=1}^{n} a_{ki}\mathbf{e}_k\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} a'_{ij}a_{ki}\right)\mathbf{e}_k$$

Como las coordenadas son únicas, $\displaystyle\sum_{i=1}^{n} a_{ki}a'_{ij} = \delta_{kj}$, donde $\delta_{kj}$ es el símbolo de Kronecker.

Pero esta suma representa al elemento $jk$ de la matriz $AA'$. Es decir, $AA' = I_n$, luego $A'$ es la inversa de $A$.

■

**Ejemplo 4.10**

(I) Consideremos la base canónica de $\mathbb{R}^3$, $\mathcal{B}_c = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ con

$$\mathbf{e}_1 = (1,0,0), \quad \mathbf{e}_2 = (0,1,0), \quad \mathbf{e}_3 = (0,0,1)$$

y el conjunto $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, con

$$\mathbf{u}_1 = \mathbf{e}_1 + \mathbf{e}_3, \quad \mathbf{u}_2 = \mathbf{e}_2, \quad \mathbf{u}_3 = \mathbf{e}_2 + \mathbf{e}_3$$

Para encontrar las ecuaciones del cambio de base de $\mathcal{B}$ a $\mathcal{B}_c$ o viceversa, primero tenemos que encontrar la relación entre los vectores de ambas bases. Dado que los vectores de la base $\mathcal{B}$ están escritos respecto de la base canónica, esto es,

$$\mathbf{u}_1 = (1,0,1), \quad \mathbf{u}_2 = (0,1,0), \quad \mathbf{u}_3 = (0,1,1)$$

obtenemos de forma inmediata la matriz del cambio de $\mathcal{B}$ a $\mathcal{B}_c$, es decir,

$$M_{\mathcal{B}}^{\mathcal{B}_c} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Para encontrar las coordenadas del vector $\mathbf{x} = 3\mathbf{u}_1 + 2\mathbf{u}_2 = (3, 2, 0)_{\mathcal{B}}$ en la base canónica hacemos uso de las ecuaciones del cambio de base, esto es $\mathbf{x}_{\mathcal{B}_c} = M_{\mathcal{B}}^{\mathcal{B}_c}\mathbf{x}_{\mathcal{B}}$. Así,

$$\mathbf{x}_{\mathcal{B}_c} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix}$$

Si ahora tenemos un vector $\mathbf{x} = (2, 1, -3)$ respecto de la base canónica, ¿cómo calcular sus coordenadas respecto de la base $\mathcal{B}$? En este caso podemos buscar escalares tales que

$$(2, 1, -3) = \alpha_1(1, 0, 1) + \alpha_2(0, 1, 0) + \alpha_3(0, 1, 1)$$

y resolver el sistema, tal y como hicimos en (II) del ejemplo 4.6; o bien usamos las ecuaciones del cambio de base del siguiente modo:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \mathbf{x}_{\mathcal{B}} = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$$

y resolvemos el sistema, resultando $\mathbf{x} = (2, 6, -5)_{\mathcal{B}}$. Nótese que el sistema anterior equivale a

$$\mathbf{x}_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$$

que corresponde al cambio de base inverso. Es decir, la matriz $M_{\mathcal{B}_c}^{\mathcal{B}}$ es $(M_{\mathcal{B}}^{\mathcal{B}_c})^{-1}$

(II) Consideremos ahora las bases

$$\mathcal{B}_1 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}, \quad \mathcal{B}_2 = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}, \quad \mathcal{B}_3 = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

donde

$$\begin{aligned} \mathbf{u}_1 &= 2\mathbf{e}_1 + \mathbf{e}_2 & \mathbf{v}_1 &= \mathbf{u}_2 + \mathbf{u}_3 \\ \mathbf{u}_2 &= \mathbf{e}_1 - \mathbf{e}_2 & \mathbf{v}_2 &= \mathbf{u}_1 - 2\mathbf{u}_2 \\ \mathbf{u}_3 &= \mathbf{e}_1 - \mathbf{e}_3 & \mathbf{v}_3 &= 2\mathbf{u}_2 - \mathbf{u}_3 \end{aligned}$$

y el vector $\mathbf{x} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$. ¿Cuáles son las coordenadas de este vector en cada una de las bases?

Observemos en primer lugar que las coordenadas de $\mathbf{x}$ en la base $\mathcal{B}_1$ son inmediatas: $\mathbf{x}_{\mathcal{B}_1} = (1, 1, 1)$. Para calcular las coordenadas en las otras dos bases haremos uso de la información proporcionada para construir las matrices de cambio de base. Dado que nos dan los vectores de la base $\mathcal{B}_2$ en función de los vectores de la base $\mathcal{B}_1$, es fácil encontrar la matriz del cambio de base de $\mathcal{B}_2$ a $\mathcal{B}_1$ (poniendo las coordenadas por columnas):

$$M_{\mathcal{B}_2}^{\mathcal{B}_1} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Y por el mismo motivo,

$$M_{\mathcal{B}_3}^{\mathcal{B}_2} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix}$$

Para obtener las coordenadas de $\mathbf{x}$ en la base $\mathcal{B}_2$ usamos la matriz del cambio de base del siguiente modo; denotando por $\mathbf{x}_{\mathcal{B}_2} = (x_1, x_2, x_3)$

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \mathbf{x}_{\mathcal{B}_2} = (1, 0 - 1)$$

Y procediendo de igual modo, si $\mathbf{x}_{\mathcal{B}_3} = (y_1, y_2, y_3)$,

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \Rightarrow \mathbf{x}_{\mathcal{B}_3} = (0, 1, 1)$$

¿Podría el lector encontrar la matriz del cambio de base de $\mathcal{B}_1$ a $\mathcal{B}_3$?

4 5

## SUBESPACIOS VECTORIALES

La última sección de este tema está dedicada a los conjuntos que más vamos a usar de ahora en adelante, los subespacios vectoriales.

**Definición 4.10**

Sea $V$ un e.v. sobre un cuerpo $\mathbb{K}$ y $W \subset V$ un subconjunto suyo. Se dice que $W$ es un *subespacio vectorial* (o variedad lineal) de $V$ si $W$ es un espacio vectorial sobre $\mathbb{K}$ con las operaciones definidas en $V$.

La definición de subespacio vectorial es un tanto confusa, pues en realidad no es más que un espacio vectorial que "vive" dentro de otro, que habitualmente denominaremos *espacio ambiente*. Puesto que las operaciones definidas en el espacio ambiente satisfacen las propiedades oportunas para ser e.v. (véase la Definición 4.1), la clave para que un subconjunto $W$ sea subespacio vectorial es que las operaciones se mantengan definidas dentro del conjunto, es decir, que la suma de vectores de $W$ sea un vector de $W$ y que el producto de un escalar por un vector de $W$ permanezca en $W$. El siguiente resultado expresa justamente esto.

**Proposición 4.9**

$W$ es un subespacio vectorial si y sólo si $\alpha\mathbf{u}+\beta\mathbf{v} \in W$, $\forall \mathbf{u}, \mathbf{v} \in W$, $\alpha$, $\beta \in \mathbb{K}$.

La demostración es inmediata, y se propone como ejercicio al lector.

**Ejemplo 4.11**

(I) El conjunto $\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}$ es un subespacio vectorial de $\mathbb{R}^2$.

Para comprobarlo, consideremos un par de elementos cualesquiera del conjunto. Dada la naturaleza del mismo, dos elementos genéricos del conjunto serán algo del estilo $(x, 0)$, $(y, 0)$. Tomemos ahora una combinación lineal arbitraria de estos dos elementos:

$$\alpha(x, 0) + \beta(y, 0) = (\alpha x + \beta y, 0)$$

Es evidente que el elemento $(\alpha x+\beta y, 0)$ pertenece al conjunto, cualesquiera que sean $\alpha$ y $\beta$; esto prueba que dicho conjunto es subespacio vectorial.

(II) El conjunto $\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 1\}$ no es subespacio vectorial de $\mathbb{R}^2$.

Si tratamos de repetir el argumento anterior considerando un par de elementos arbitrarios del conjunto, $(x, 1)$, $(y, 1)$, ahora vemos que una combinación lineal de los mismos,

$$\alpha(x, 1) + \beta(y, 1) = (\alpha x + \beta y, \alpha + \beta)$$

no es, en general, un elemento del conjunto. En concreto, si tomamos valores $\alpha$ y $\beta$ tales que $\alpha + \beta \neq 1$, la combinación lineal resultante no pertenece al conjunto, y por tanto éste no es subespacio vectorial.

(III) $\mathbb{P}^n_{\mathbb{R}}[x]$ es un subespacio vectorial de $\mathbb{P}_{\mathbb{R}}[x]$.

(IV) $\mathbb{P}_{\mathbb{R}}[x]$ es un subespacio vectorial de $\mathcal{C}(\mathbb{R})$.

(V) $\{\mathbf{0}\}$ es un subespacio vectorial de cualquier e.v.

**Definición 4.11**

Sean $\mathbf{u}_1, \ldots, \mathbf{u}_k$ vectores de un e.v. $V$. Se define el conjunto

$$W = L(\mathbf{u}_1, \ldots, \mathbf{u}_k) \equiv \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle = \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{u}_i : \alpha_i \in \mathbb{K} \right\}$$

es decir, el conjunto de todas las posibles combinaciones lineales que se pueden hacer con los vectores dados. Este conjunto se denomina *subespacio engendrado por* $\mathbf{u}_1, \ldots, \mathbf{u}_k$ y a tales vectores se les denomina sistema generador de $W$.

**Proposición 4.10**

$W = L(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ es un subespacio vectorial de $V$.

*Demostración:*
Para la prueba usaremos la Proposición 4.9. Sean $\mathbf{u}, \mathbf{v} \in W$, y sean $\alpha, \beta \in \mathbb{K}$.

Veamos que $\alpha\mathbf{u}+\beta\mathbf{v}$ está en $W$. En efecto,

$$\alpha\mathbf{u}+\beta\mathbf{v}=\alpha\left(\sum_{i=1}^{k}\alpha_i\mathbf{u}_i\right)+\beta\left(\sum_{i=1}^{k}\beta_i\mathbf{u}_i\right)=\sum_{i=1}^{k}(\alpha\alpha_i+\beta\beta_i)\mathbf{u}_i\in W$$

■

Nótese que denominamos sistema generador a un conjunto de vectores de $V$ que no es sistema generador de $V$, sino de un subespacio. Pero el nombre es consistente con el significado de sistema generador: un conjunto que genera el subespacio.

**Ejemplo 4.12**

(I) En $\mathbb{R}^3$, estudiar el espacio vectorial generado por los vectores $\mathbf{u}_1=(1,0,1)$, $\mathbf{u}_2=(-1,1,0)$ y $\mathbf{u}_3=(1,1,2)$.

Puesto que estos tres vectores conforman un sistema generador de $L=\langle\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\rangle$, en primer lugar estudiamos la dependencia lineal entre los mismos. Así,

$$\text{rango}\begin{pmatrix}1&0&1\\-1&1&0\\1&1&2\end{pmatrix}=2\quad\text{pues}\quad\begin{vmatrix}1&0\\-1&1\end{vmatrix}\neq 0\quad\text{y}\quad\begin{vmatrix}1&0&1\\-1&1&0\\1&1&2\end{vmatrix}=0$$

lo que significa que los vectores $\mathbf{u}_1$ y $\mathbf{u}_2$ son l.i y que $\mathbf{u}_3$ es combinación lineal de ellos. Es decir, $\mathbf{u}_1$ y $\mathbf{u}_2$ generan todo $L$, y puesto que son l.i. forman una base de $L$. La dimensión de $L$ será por tanto 2 (el número de elementos de una base), y $\langle\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\rangle=\langle\mathbf{u}_1,\mathbf{u}_2\rangle$.

Es importante que el lector asimile correctamente estos conceptos que pueden ser un tanto confusos al principio. Obsérvese que el conjunto $\{\mathbf{u}_1,\mathbf{u}_2\}$ es base de $L$ pero no es base de $\mathbb{R}^3$ (pues genera $L$, pero no $\mathbb{R}^3$). Podemos incluso usar coordenadas en $L$; por ejemplo, el vector $\mathbf{u}_3=2\mathbf{u}_1+\mathbf{u}_2$, tiene coordenadas $(2,1)_L$ respecto de la base dada en $L$. Aunque pueda parecer extraño que sólo haya dos coordenadas, hemos de tener en cuenta que el subespacio $L$ es de dimensión 2.

(II) Sea ahora $L=L(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$, con $\mathbf{u}_1=(1,0,1)$, $\mathbf{u}_2=(1,1,1)$ y $\mathbf{u}_3=(0,1,2)$. Nuevamente estudiamos el rango del conjunto de vectores dado,

$$\text{rango}\begin{pmatrix}1&0&1\\1&1&1\\0&1&2\end{pmatrix}=3$$

lo que indica que los tres vectores son l.i. Como estamos en $\mathbb{R}^3$, cuya dimensión es 3, estos vectores forman una base (recuérdese el Corolario 4.5). Es decir, este conjunto de vectores es sistema generador de todo el espacio $\mathbb{R}^3$, luego $L(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \mathbb{R}^3$.

Hemos visto como construir subespacios a partir de un conjunto de vectores. El siguiente resultado nos muestra otra forma habitual de definir subespacios.

**Teorema 4.7**

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$, con $\operatorname{rango}(A) = r$. El conjunto de soluciones del sistema homogéneo $A\mathbf{x} = \mathbf{0}$ es un subespacio vectorial generado por cualesquiera $k = n - r$ soluciones linealmente independientes del sistema.

*Demostración:*
La idea de la demostración consiste en construir un conjunto de soluciones independientes que genere a todas las demás; de este modo, el conjunto de soluciones será un subespacio vectorial.

Por simplicidad en la notación supondremos que un menor de orden $r$ distinto de cero está formado por las primeras $r$ filas y las primeras $r$ columnas. Es decir,

$$\begin{vmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} \end{vmatrix} \neq 0 \tag{4.7}$$

Como ya vimos en el Teorema 4.1, las restantes filas son combinación lineal de las $r$ primeras, es decir, el sistema

$$\left.\begin{array}{c} a_{11}x_1 + \cdots + a_{1r}x_r + \cdots + a_{1n}x_n = 0 \\ \vdots \\ a_{r1}x_1 + \cdots + a_{rr}x_r + \cdots + a_{rn}x_n = 0 \end{array}\right\}$$

tiene las mismas soluciones que el sistema original. Entonces, pasando las incógnitas $x_{r+1}, \dots, x_n$ al segundo miembro,

$$\left.\begin{array}{ccc} a_{11}x_1 + \cdots + a_{1r}x_r & = & -a_{1,r+1}x_{r+1} - \cdots - a_{1n}x_n \\ \vdots & \vdots & \vdots \\ a_{r1}x_1 + \cdots + a_{rr}x_r & = & -a_{r,r+1}x_{r+1} - \cdots - a_{rn}x_n \end{array}\right\} \tag{4.8}$$

este sistema es un sistema de Cramer cuya matriz de los coeficientes coincide con el menor básico dado en (4.7).

Denotemos por $B$ a la matriz que corresponde a este menor, y consideremos los $n-r$ vectores de $\mathbb{R}^r$ siguientes

$$\mathbf{b}_{r+1} = \begin{pmatrix} a_{1,r+1} \\ \vdots \\ a_{r,r+1} \end{pmatrix} \quad \cdots \quad \mathbf{b}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{rn} \end{pmatrix}$$

Es fácil comprobar que el sistema (4.8) se puede escribir como

$$B \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = -\mathbf{b}_{r+1}x_{r+1} - \cdots - \mathbf{b}_n x_n$$

o lo que es lo mismo,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = -B^{-1}\mathbf{b}_{r+1}x_{r+1} - \cdots - B^{-1}\mathbf{b}_n x_n \tag{4.9}$$

(recordemos que existe la inversa de $B$ gracias a (4.7)).

Consideremos ahora el vector de $\mathbb{R}^n$

$$\mathbf{s}_1 = \begin{pmatrix} -B^{-1}\cdot \mathbf{b}_{r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \}\, r \text{ filas} \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} n-r \text{ filas} \end{matrix}$$

es decir, hemos resuelto (4.9) para $x_{r+1} = 1$, $x_{r+2} = \cdots = x_n = 0$, y análogamente los vectores

$$\mathbf{s}_j = \begin{pmatrix} -B^{-1}\cdot \mathbf{b}_{r+j} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \longleftarrow \text{Fila } r+j \\ \\ \\ \end{matrix}$$

para $j = 2, \dots, n-r$, obtenidos al resolver (4.9) para $x_{r+1} = \cdots = x_{r+j-1} = x_{r+j+1} = \cdots = x_n = 0$ y $x_{r+j} = 1$.

Entonces,

(I) Es evidente por construcción que los vectores $\mathbf{s}_1, \dots, \mathbf{s}_{n-r}$ son solución del sistema original.

(II) $\{\mathbf{s}_1, \dots, \mathbf{s}_{n-r}\}$ es l.i.

Para ver esto basta observar la matriz obtenida poniendo dichos vectores por filas. Las $n-r$ últimas filas de esta matriz corresponden exactamente a $I_{n-r}$, luego ahí tenemos un menor de orden $n-r$ distinto de cero.

(III) $\{\mathbf{s}_1, \dots, \mathbf{s}_{n-r}\}$ es un sistema generador de las soluciones del sistema original.

Para probar esto consideremos una solución del sistema, $\mathbf{p} = (p_1, \dots, p_n)$. La construcción anterior nos lleva a que $\mathbf{p}$ satisface

$$B \begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix} = -\mathbf{b}_{r+1} p_{r+1} - \cdots - \mathbf{b}_n p_n$$

pero teniendo en cuenta cómo son las soluciones $\mathbf{s}_1, \dots, \mathbf{s}_{\mathbf{n-r}}$ es fácil comprobar que

$$\mathbf{p} = p_{r+1}\mathbf{s}_1 + \cdots + p_n \mathbf{s}_{n-r}$$

es decir, cualquier solución del sistema es combinación lineal de los vectores $\mathbf{s}_1, \dots, \mathbf{s}_{n-r}$.

Esto demuestra que $\{\mathbf{s}_1, \dots, \mathbf{s}_{n-r}\}$ es una base de las soluciones del sistema homogéneo dado, y de aquí se sigue el resultado.

■

**Nota 4.7**

Como consecuencia de este resultado, las soluciones de un sistema lineal homogéneo conforman un subespacio vectorial de dimensión $n-\text{rango}(A)$, siendo $n$ el número de incógnitas y $A$ la matriz de los coeficientes. Obsérvese que la dimensión del subespacio coincide con el número de grados de libertad del sistema (número de incógnitas menos número de ecuaciones relevantes).

**Ejemplo 4.13**

Encontrar una base del conjunto de soluciones del sistema homogéneo:

$$\left.\begin{array}{rcl} x_2 + x_3 + x_4 + x_5 & = & 0 \\ x_1 - x_2 & = & 0 \\ x_1 + x_3 + x_4 + x_5 & = & 0 \\ x_1 + x_2 + x_3 + x_4 + x_5 & = & 0 \end{array}\right\}$$

En primer lugar estudiamos el rango de la matriz asociada a fin de conocer la dimensión del subespacio generado por las soluciones.

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Obsérvese que las columnas tercera, cuarta y quinta son iguales, por lo que las dos últimas no influyen en el rango. De manera que el rango no puede ser superior a tres. Un simple cálculo demuestra que el determinante

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \neq 0$$

de lo que deducimos que $\text{rango}(A) = 3$. La dimensión del espacio generado será por tanto $5 - 3 = 2$. Para encontrar una base de las soluciones, pasamos al segundo miembro las incógnitas correspondientes a las columnas que no forman parte del menor básico encontrado (al igual que en (4.8)); en este caso $x_4$ y $x_5$, y eliminamos las ecuaciones que no forman parte del menor básico, por ser redundantes. Esto es,

$$\left.\begin{array}{rcl} x_2 + x_3 & = & -x_4 - x_5 \\ x_1 - x_2 & = & 0 \\ x_1 + x_2 + x_3 & = & -x_4 - x_5 \end{array}\right\}$$

Ahora resolvemos el sistema usando la técnica empleada en la demostración del teorema anterior. Es decir, para

$$x_4 = 1,\ x_5 = 0 \longrightarrow \left.\begin{array}{rcl} x_2 + x_3 & = & -1 \\ x_1 - x_2 & = & 0 \\ x_1 + x_2 + x_3 & = & -1 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = 0 \\ x_2 = 0 \\ x_3 = -1 \end{array}$$

luego la solución es $(0,0,-1,1,0)$, y para

$$x_4 = 0,\ x_5 = 1 \longrightarrow \left.\begin{array}{rcr} x_2 + x_3 & = & -1 \\ x_1 - x_2 & = & 0 \\ x_1 + x_2 + x_3 & = & -1 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = 0 \\ x_2 = 0 \\ x_3 = -1 \end{array}$$

la solución es $(0,0,-1,0,1)$. Por tanto, las soluciones de este sistema coinciden con el subespacio de dimensión 2:

$$L = L((0,0,-1,1,0),(0,0,-1,0,1))$$

**Ejemplo 4.14**

En $\mathbb{R}^3$ consideramos el sistema

$$\left.\begin{array}{rcr} x_1 + x_2 & = & 0 \\ x_2 & = & 0 \end{array}\right\}$$

Obtener una base del subespacio formado por sus soluciones. ¿Pertenece el vector $(-1,1,0)$ a dicho subespacio?

En este ejemplo es esencial prestar atención al espacio en el que se encuentran las soluciones del sistema, y que viene resaltado en el enunciado: $\mathbb{R}^3$. De no haber sido especificado, habríamos supuesto, de forma natural, que el espacio correspondería a $\mathbb{R}^2$, pues vemos solo dos incógnitas en el sistema. Al tratarse de un sistema en $\mathbb{R}^3$, hemos de considerar tres incógnitas (a pesar de que no las veamos). La matriz del sistema es por tanto

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

cuyo rango es dos, y de donde deducimos que la dimensión del espacio de soluciones es $3 - 2 = 1$. Para obtener una base hemos de resolver el sistema usando un parámetro al que daremos un único valor (igual a 1). Al igual que hemos hecho en el ejemplo anterior, los parámetros deben ser las incógnitas correspondientes a las columnas de la matriz que no forman parte del menor que proporciona el rango de la misma, es decir, en este caso, $x_3$:

$$x_3 = 1 \longrightarrow \left.\begin{array}{rcr} x_1 + x_2 & = & 0 \\ x_2 & = & 0 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = 0 \\ x_2 = 0 \end{array}$$

es decir, una base de soluciones está formada por el vector $(0, 0, 1)$.

En cuanto a la pertenencia del vector $(-1, 1, 0)$, la forma más sencilla de comprobarlo consiste en ver si el vector $(-1, 1, 0)$ verifica las ecuaciones que definen el subespacio, esto es, $x_1 + x_2 = 0$, lo cual se cumple, y $x_2 = 0$, que no se cumple. Puesto que no verifica todas las ecuaciones, este vector no éstá en el subespacio.

También podríamos haber razonado comprobando que el vector $(-1, 1, 0)$ es independiente con la base que hemos obtenido, de modo que no está en dicho subespacio.

**Teorema 4.8**

Recíprocamente, todo subespacio vectorial de $\mathbb{K}^n$ de dimensión $k$ puede determinarse a través de las soluciones de un sistema lineal homogéneo.

*Demostración:*

Sea $L = L(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, con $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ un conjunto l.i. donde denotamos por $\mathbf{w}_i = (a_{i1}, \ldots, a_{in})$, $1 \leq i \leq k$.

Consiremos ahora $\mathbf{x}$ es un vector de $L$ de coordenadas $(x_1, \ldots, x_n)$. La pertenecia de $\mathbf{x}$ al subespacio $L$ significa que $\mathbf{x}$ es combinación lineal de los vectores $\mathbf{w}_i$, luego

$$\text{rango}\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} = k = \text{rango}\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \\ x_1 & \cdots & x_n \end{pmatrix}$$

Es decir, todos los menores de orden $k+1$ de la última matriz tienen que ser necesariamente nulos. Así pues, bastará orlar un menor básico distinto de cero (que por comodidad suponemos que se tiene en las $k$ primeras columnas) con la fila $k+1$ y las restantes columnas, o sea,

$$\begin{vmatrix} a_{11} & \cdots & a_{1k} & a_{1,k+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k1} & \cdots & a_{kk} & a_{k,k+1} \\ x_1 & \cdots & x_k & x_{k+1} \end{vmatrix} = 0, \quad \begin{vmatrix} a_{11} & \cdots & a_{1k} & a_{1,k+2} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k1} & \cdots & a_{kk} & a_{k,k+2} \\ x_1 & \cdots & x_k & x_{k+2} \end{vmatrix} = 0, \quad \ldots$$

$$\begin{vmatrix} a_{11} & \cdots & a_{1k} & a_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k1} & \cdots & a_{kk} & a_{kn} \\ x_1 & \cdots & x_k & x_n \end{vmatrix} = 0$$

Cada uno de estos determinantes proporciona una ecuación lineal y por tanto se genera de este modo un sistema lineal homogéneo de $n-k$ ecuaciones, cuyas soluciones son vectores de $L$.

■

A un sistema de ecuaciones homogéneo cuyas soluciones determinan un subespacio se le denomina *sistema de ecuaciones implícitas* de ese subespacio.

**Ejemplo 4.15**

Consideremos el subespacio $L = \langle(1,1,1),(0,1,0)\rangle$ y encontremos una base y un sistema de ecuaciones implícitas del mismo.

Dado que

$$\text{rango}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} = 2$$

los vectores que generan el subespacio son l.i. y por tanto forman una base. Para encontrar un sistema de ecuaciones implícitas debemos pedir que

$$\text{rango}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ x_1 & x_2 & x_3 \end{pmatrix} = 2 \Rightarrow \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ x_1 & x_2 & x_3 \end{vmatrix} = 0 \Rightarrow x_3 - x_1 = 0$$

Así pues, el sistema formado por la ecuación $x_1 - x_3 = 0$ es un sistema de ecuaciones implícitas de $L$.

**Ejemplo 4.16**

Procedamos de forma análoga con el subespacio $L = L(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$, donde

$$\mathbf{w}_1 = (1,1,1,1), \quad \mathbf{w}_2 = (-3,2,0,1), \quad \mathbf{w}_3 = (-2,3,1,2)$$

Sabemos que

$$\dim L = \text{rango}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & 2 & 0 & 1 \\ -2 & 3 & 1 & 2 \end{pmatrix} = 2 \quad \text{con} \quad \text{rango}\begin{vmatrix} 1 & 1 \\ -3 & 2 \end{vmatrix} \neq 0$$

luego una base de $L$ vendrá dada por los vectores $\mathbf{w}_1$ y $\mathbf{w}_2$. Para encontrar un sistema de ecuaciones implícitas procedemos como en el caso anterior:

$$\text{rango}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & 2 & 0 & 1 \\ x_1 & x_2 & x_3 & x_4 \end{pmatrix} = 2 \Rightarrow \begin{cases} \begin{vmatrix} 1 & 1 & 1 \\ -3 & 2 & 0 \\ x_1 & x_2 & x_3 \end{vmatrix} = 0, \\ \begin{vmatrix} 1 & 1 & 1 \\ -3 & 0 & 1 \\ x_1 & x_3 & x_4 \end{vmatrix} = 0 \end{cases}$$

de donde se obtiene el sistema de ecuaciones

$$\begin{cases} -2x_1 - 3x_2 + 5x_3 = 0 \\ x_1 - 4x_3 + 3x_4 = 0 \end{cases}$$

Nótese que hay más menores de orden tres que no hemos considerado, (por ejemplo, el formado por las tres últimas columnas de la matriz) que darían lugar a otra ecuación distinta. Sin embargo, el método del orlado garantiza que esta ecuación resultante es dependiente de las anteriores, y por tanto no es necesario considerarla.

**Ejemplo 4.17**

Encontrar la dimensión y una base del subespacio

$$M = \{p(x) \in \mathbb{P}^2_{\mathbb{R}}[x] : p(1) = 0\}$$

En este ejemplo nos enfrentamos con una forma diferente de definir un subespacio vectorial. En lugar de proporcionarnos un sistema de generadores (como en el ejemplo 4.12) o un sistema de ecuaciones homogéneo (como en el ejemplo 4.13) nos dan una descripción de los elementos que forman parte del conjunto, a partir de la cual hemos de encontrar su dimensión y una base.

Para ello procederemos del siguiente modo: escogemos un elemento genérico del espacio que contiene al conjunto, en este caso $\mathbb{P}^2_{\mathbb{R}}[x]$, para el que usaremos las incógnitas $x_i$ como coordenadas canónicas, es decir, tomamos

$$p(x) = x_1 + x_2 \cdot x + x_3 \cdot x^2$$

y a continuación imponemos sobre este elemento la descripción que define a los elementos de nuestro conjunto. Esto es,

$$p(x) \in M \Rightarrow p(1) = 0 \Rightarrow x_1 + x_2 + x_3 = 0$$

De este modo se obtiene un sistema de ecuaciones homogéneo (en este caso solo una ecuación) para el que llevamos a cabo un tratamiento como en el ejemplo 4.13. Puesto que el rango de la matriz de este sistema

$$A = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$$

es uno y $\dim(\mathbb{P}^2_{\mathbb{R}}[x]) = 3$, entonces $\dim(M) = 3 - 1 = 2$ y resolvemos la ecuación con dos parámetros:

$$\begin{aligned} x_2 = 1,\ x_3 = 0 &\longrightarrow x_1 = -1 \\ x_2 = 0,\ x_3 = 1 &\longrightarrow x_1 = -1 \end{aligned}$$

luego las soluciones son $(-1, 1, 0)$ y $(-1, 0, 1)$. Nótese que estas soluciones corresponden a las coordenadas respecto de la base canónica de $\mathbb{P}^2_{\mathbb{R}}[x]$, por lo tanto, una base de $M$ está formada por $\{-1 + x, -1 + x^2\}$.

### 4.5.1 Operaciones con subespacios

En lo que resta de tema vamos a trabajar con más de un subespacio vectorial del mismo e.v. Comenzamos viendo una propiedad interesante de los subespacios vectoriales.

**Teorema 4.9**

Sean $L_1$ y $L_2$ dos subespacios vectoriales de $V$. Si $L_1 \subset L_2$ y $\dim(L_1) = \dim(L_2)$ entonces $L_1 = L_2$.

*Demostración:*

Puesto que $L_1 \subset L_2$, sólo hay que probar la inclusión contraria.[7] Sea $\mathbf{x} \in L_2$, y $\mathcal{B}$ una base de $L_1$ formada por $\dim(L_1) = k$ vectores. Como $L_1 \subset L_2$, el conjunto $\mathcal{B}$ es l.i. en $L_2$. Como además $L_2$ también tiene dimensión igual a $k$, $\mathcal{B}$ es un conjunto de $k$ vectores l.i. en un espacio de dimensión $k$, es decir es una base de $L_2$ (véase el Corolario 4.5).

---

[7] Cuando queremos probar que dos conjuntos $A$ y $B$ son iguales, se suele proceder probando la doble inclusión, es decir $A \subset B$ y $B \subset A$ (véase el apéndice A).

Al ser base de $L_2$, $\mathbf{x}$ se escribe como combinación lineal de $\mathcal{B}$, que a su vez está también en $L_1$, por tanto $\mathbf{x} \in L_1$. De la arbitrariedad de $\mathbf{x}$ se sigue la igualdad entre los espacios.

■

Nótese que la clave en la demostración anterior está en probar que la base de $L_1$ también lo es de $L_2$. En definitiva, si un conjunto es base de dos espacios, entonces ambos espacios son iguales (de hecho basta con que sea sistema generador).

**Definición 4.12**

Dados dos subespacios $L_1$, $L_2$ de un e.v. $V$ se define la *suma* de $L_1$ y $L_2$ por

$$L_1 + L_2 = \{\mathbf{u}_1 + \mathbf{u}_2 : \mathbf{u}_1 \in L_1,\ \mathbf{u}_2 \in L_2\}$$

Igualmente definimos la *intersección* de $L_1$ y $L_2$ por

$$L_1 \cap L_2 = \{\mathbf{u} : \mathbf{u} \in L_1,\ \mathbf{u} \in L_2\}$$

**Teorema 4.10**

Si $L_1$ y $L_2$ son subespacios vectoriales, entonces $L_1 + L_2$ y $L_1 \cap L_2$ también son subespacios vectoriales.

*Demostración:*
Usaremos la Proposición 4.9. Si $\mathbf{u}$ y $\mathbf{v} \in L_1 + L_2$ entonces

$$\begin{aligned} \mathbf{u} &= \mathbf{u}_1 + \mathbf{u}_2, \quad \text{con } \mathbf{u}_1 \in L_1,\ \mathbf{u}_2 \in L_2 \\ \mathbf{v} &= \mathbf{v}_1 + \mathbf{v}_2, \quad \text{con } \mathbf{v}_1 \in L_1,\ \mathbf{v}_2 \in L_2 \end{aligned}$$

Entonces,

$$\alpha\mathbf{u} + \beta\mathbf{v} = \alpha(\mathbf{u}_1 + \mathbf{u}_2) + \beta(\mathbf{v}_1 + \mathbf{v}_2) = (\alpha\mathbf{u}_1 + \beta\mathbf{v}_1) + (\alpha\mathbf{u}_2 + \beta\mathbf{v}_2)$$

con

$$(\alpha\mathbf{u}_1 + \beta\mathbf{v}_1) \in L_1 \quad \text{y} \quad (\alpha\mathbf{u}_2 + \beta\mathbf{v}_2) \in L_2$$

Del mismo modo, si $\mathbf{u}, \mathbf{v} \in L_1 \cap L_2$ entonces

$$\left.\begin{array}{l} \mathbf{u} \in L_1,\ \mathbf{v} \in L_1 \Rightarrow \alpha\mathbf{u} + \beta\mathbf{v} \in L_1 \\ \mathbf{u} \in L_2,\ \mathbf{v} \in L_2 \Rightarrow \alpha\mathbf{u} + \beta\mathbf{v} \in L_2 \end{array}\right\} \Rightarrow \alpha\mathbf{u} + \beta\mathbf{v} \in L_1 \cap L_2$$

■

El siguiente resultado nos dice cómo obtener la suma y la intersección de subespacios a partir de las bases o los sistemas de ecuaciones implícitas.

**Teorema 4.11**

Sean $L_1$ y $L_2$ dos subespacios de un e.v. $V$. Se verifica:

(I) Si $\mathcal{B}_1$ es una base de $L_1$ y $\mathcal{B}_2$ es una base de $L_2$, entonces $L_1 + L_2 = L(\mathcal{B}_1 \cup \mathcal{B}_2)$, esto es, $\mathcal{B}_1 \cup \mathcal{B}_2$ es un sistema generador de $L_1 + L_2$.

(II) Si $A\mathbf{x} = \mathbf{0}$ es un sistema de ecuaciones implícitas de $L_1$ y $B\mathbf{x} = \mathbf{0}$ es un sistema de ecuaciones implícitas de $L_2$, entonces,

$$\left.\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ B\mathbf{x} &= \mathbf{0} \end{aligned}\right\}$$

es un sistema de ecuaciones implícitas de $L_1 \cap L_2$.

La demostración se deja como ejercicio para el lector (ejercicio 28).

**Teorema 4.12 (Fórmula de la dimensión)**

Si $L_1$ y $L_2$ son subespacios de un e.v. $V$ se verifica:

$$\dim(L_1) + \dim(L_2) = \dim(L_1 + L_2) + \dim(L_1 \cap L_2)$$

*Demostración:*
Sea $\{\mathbf{e}_1, \dots, \mathbf{e}_l\}$ una base de $L_1 \cap L_2$. Como $L_1 \cap L_2 \subset L_1, L_2$ podemos encontrar vectores $\mathbf{f}_1, \dots \mathbf{f}_k$, $\mathbf{g}_1, \dots \mathbf{g}_m$, tales que

$$\begin{aligned} &\{\mathbf{e}_1, \dots, \mathbf{e}_l, \mathbf{f}_1, \dots, \mathbf{f}_k\} \quad \text{es base de } L_1 \\ &\{\mathbf{e}_1, \dots, \mathbf{e}_l, \mathbf{g}_1, \dots, \mathbf{g}_m\} \quad \text{es base de } L_2 \end{aligned}$$

Está claro que si probamos que el conjunto

$$\mathcal{S} = \{\mathbf{e}_1, \dots, \mathbf{e}_l, \mathbf{f}_1, \dots, \mathbf{f}_k, \mathbf{g}_1, \dots, \mathbf{g}_m\}$$

es una base de $L_1 + L_2$ habremos terminado la prueba.

Es evidente, usando (I) del Teorema 4.11, que $\mathcal{S}$ es un sistema generador de $L_1 + L_2$. Veamos que también es l.i. Consideremos una combinación lineal igualada al vector nulo:

$$\alpha_1\mathbf{e}_1 + \cdots + \alpha_l\mathbf{e}_l + \beta_1\mathbf{f}_1 + \cdots + \beta_k\mathbf{f}_k + \gamma_1\mathbf{g}_1 + \cdots + \gamma_m\mathbf{g}_m = \mathbf{0} \qquad (4.10)$$

Denotemos por $\mathbf{v} = \alpha_1\mathbf{e}_1 + \cdots + \alpha_l\mathbf{e}_l + \beta_1\mathbf{f}_1 + \cdots + \beta_k\mathbf{f}_k$. Está claro que $\mathbf{v} \in L_1$. Por otra parte, podemos escribir

$$\mathbf{v} = -\gamma_1\mathbf{g}_1 - \cdots - \gamma_m\mathbf{g}_m$$

luego $\mathbf{v} \in L_2$. De modo que $\mathbf{v} \in L_1 \cap L_2$. Como las coordenadas respecto de una base son únicas,

$$\beta_1 = \cdots = \beta_k = \gamma_1 = \cdots \gamma_m = 0 \qquad (4.11)$$

Finalmente, como $\mathbf{e}_1, \ldots, \mathbf{e}_l$ forman una base, de (4.10) y (4.11) deducimos que $\alpha_1 = \cdots = \alpha_l = 0$.

■

Finalizamos la sección con el concepto de *suma directa* de subespacios.

**Definición 4.13**

Un e.v. $V$ es *suma directa* de dos subespacios $L_1$ y $L_2$ si y sólo si

$$L_1 + L_2 = V \quad \text{y} \quad L_1 \cap L_2 = \{\mathbf{0}\}$$

Se notará $V = L_1 \oplus L_2$.

**Teorema 4.13**

Son equivalentes:

(I) $V = L_1 \oplus L_2$.

(II) $\forall\mathbf{v} \in V$, $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, con $\mathbf{v}_1 \in L_1$, $\mathbf{v}_2 \in L_2$, únicos.

*Demostración:*
Probemos en primer lugar (I) ⇒ (II).

Puesto que $V = L_1 + L_2$, sólo es necesario probar la unicidad de la descomposición. En efecto, supongamos que

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{u}_1 + \mathbf{u}_2$$

Entonces, $\mathbf{v}_1 - \mathbf{u}_1 = \mathbf{u}_2 - \mathbf{v}_2$ con $\mathbf{v}_1 - \mathbf{u}_1 \in L_1$ y $\mathbf{u}_2 - \mathbf{v}_2 \in L_2$. Es decir,

$$\mathbf{v}_1 - \mathbf{u}_1,\ \mathbf{u}_2 - \mathbf{v}_2 \in L_1 \cap L_2 = \{\mathbf{0}\}$$

De aquí se tiene que $\mathbf{v}_1 = \mathbf{u}_1$ y $\mathbf{v}_2 = \mathbf{u}_2$.

Veamos la implicación contraria. Está claro por la definición de suma de subespacios (Definición 4.12) que $V = L_1 + L_2$. Por otra parte, si $\mathbf{v} \in L_1 \cap L_2$ podemos escribir

$$\mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v}$$

Ahora bien, como la descomposición es única, sólo existe la posibilidad de que $\mathbf{v} = \mathbf{0}$; luego $L_1 \cap L_2 = \{\mathbf{0}\}$.

■

**Ejemplo 4.18**

Considérense los subespacios de $\mathbb{R}^4$

$$L_1 \equiv \left\{ \begin{array}{l} x_1 + x_2 + x_3 + x_4 = 0 \\ x_3 + x_4 = 0 \end{array} \right.$$

y

$$L_2 = \langle (1,0,0,0), (0,0,0,1), (3,0,0,7) \rangle$$

Comprobemos que $\mathbb{R}^4 = L_1 \oplus L_2$.

Calculemos una base de cada uno de los subespacios dados. Para $L_1$ hemos de resolver el sistema tal y como vimos en el ejemplo 4.13. La matriz del sistema es

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

cuyo rango es claramente dos, por tanto $\dim(L_1) = 4 - 2 = 2$. Un menor distinto de cero está formado por las columnas dos y tres, lo que nos indica que debemos resolver el sistema con los parámetros $x_1$ y $x_4$:

$$\left. \begin{array}{r} x_2 + x_3 = -x_1 - x_4 \\ x_3 = -x_4 \end{array} \right\} \Rightarrow \begin{array}{l} x_1 = 1,\ x_4 = 0 \Rightarrow x_2 = -1,\ x_3 = 0 \to (1,-1,0,0) \\ x_1 = 0,\ x_4 = 1 \Rightarrow x_2 = 0,\ x_3 = -1 \to (0,0,-1,1) \end{array}$$

Luego una base de $L_1$ es $\{(1,-1,0,0), (0,0,-1,1)\}$

Por su parte, $\{(1,0,0,0), (0,0,0,1), (3,0,0,7)\}$ es un sistema generador de $L_2$, y dado que

$$\text{rango} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 7 \end{pmatrix} = 2$$

una base de $L_2$ es $\{(1,0,0,0),(0,0,0,1)\}$, y por tanto $\dim(L_2)=2$.

Calculemos ahora $L_1+L_2$. Para ello, simplemente consideramos el conjunto formado por la base de $L_1$ y la base de $L_2$ y estudiamos si son o no l.i.:

$$\text{rango}\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4$$

luego forman base. Así pues $\dim(L_1+L_2)=4$ y dado que estamos en $\mathbb{R}^4$ se tiene que $L_1+L_2=\mathbb{R}^4$ (véase el Teorema 4.9).

Por último, estudiemos $L_1\cap L_2$. Para identificar este espacio necesitamos juntar los sistemas de ecuaciones implícitas de ambos subespacios. Para $L_1$ ya lo tenemos, y para $L_2$ procedemos como en el ejemplo 4.15:

$$\text{rango}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ x_1 & x_2 & x_3 & x_4 \end{pmatrix} = 2 \Rightarrow \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ x_1 & x_2 & x_4 \end{vmatrix} = 0, \qquad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ x_1 & x_3 & x_4 \end{vmatrix} = 0$$

que resulta:

$$x_2 = 0, \quad x_3 = 0$$

Por tanto, un sistema de ecuaciones implícitas de $L_1\cap L_2$ vendrá dado por

$$L_1\cap L_2 \equiv \begin{cases} x_1+x_2+x_3+x_4=0 \\ x_3+x_4=0 \\ x_2=0 \\ x_3=0 \end{cases}$$

Si miramos el rango de la matriz de coeficientes del sistema observamos que es 4, por tanto $\dim(L_1\cap L_2)=0$, y eso significa que $L_1\cap L_2=\{\mathbf{0}\}$.

En realidad si hubiéramos usado la fórmula de la dimensión (Teorema 4.12) podríamos haber llegado a esta conclusión mucho antes, pues sabiendo que $\dim(L_1)=2$, $\dim(L_2)=2$ y que $\dim(L_1+L_2)=4$ deducimos de inmediato que $\dim(L_1\cap L_2)$ tiene que ser 0, y por tanto es el espacio nulo. Dado que se cumplen las condiciones de la Definición 4.13 se tiene que $\mathbb{R}^4=L_1\oplus L_2$.

Finalmente, veamos como el concepto de suma e intersección entre dos

subespacios se puede generalizar a más subespacios.

**Definición 4.14**

Dados $L_1, \dots, L_n$ subespacios de un e.v. $V$, se definen

$$\sum_{i=1}^{n} L_i = \left\{ \sum_{i=1}^{n} \mathbf{v}_i : \mathbf{v}_i \in L_i \right\}$$
$$\bigcap_{i=1}^{n} L_i = \{\mathbf{v} \in V : \mathbf{v} \in L_i,\ \forall i\}$$

En este caso la suma directa de $n$ subespacios viene dada por

$$V = \bigoplus_{i=1}^{n} L_i \iff V = \sum_{i=1}^{n} L_i, \quad L_i \cap \left( \sum_{k \neq i} L_k \right) = \{\mathbf{0}\},\ i = 1, \dots, n$$

4 6

## CÁLCULO CON PYTHON

Como el lector habrá podido comprobar, los cálculos necesarios para este tema se reducen básicamente a resolver sistemas homogéneos y calcular rangos, lo cual ya se ha hecho en los temas anteriores. No obstante, si bien antes trabajamos directamente con matrices, en este tema es más frecuente trabajar con vectores, que aunque pueden ser considerados como matrices fila o columna, se perciben más como arreglos. Así pues, en esta sección insistiremos un poco más en el manejo de estos elementos.

```
>>> from numpy import array,dot
>>> a=array(range(6))
>>> a
array([0, 1, 2, 3, 4, 5])
>>> len(a)
6
>>> a.shape
(6,)
>>> b=a.reshape(2,3)
>>> b
array([[0, 1, 2],
       [3, 4, 5]])
>>> b.shape
(2, 3)
>>> c=a.reshape(6,1)
```

```
16 >>> c
17 array([[0],
18        [1],
19        [2],
20        [3],
21        [4],
22        [5]])
23 >>> b.T
24 array([[0, 3],
25        [1, 4],
26        [2, 5]])
27 >>> b*b.T
28 Traceback (most recent call last):
29   File "<stdin>", line 1, in <module>
30 ValueError: shape mismatch: objects cannot be broadcast to a
       single shape
31 >>> b*b
32 array([[ 0,  1,  4],
33        [ 9, 16, 25]])
34 >>> dot(b,b.T)
35 array([[ 5, 14],
36        [14, 50]])
37 >>> dot(b.T,b)
38 array([[ 9, 12, 15],
39        [12, 17, 22],
40        [15, 22, 29]])
41 >>> dot(a,c)
42 array([55])
43 >>> dot(c,a)
44 Traceback (most recent call last):
45   File "<stdin>", line 1, in <module>
46 ValueError: objects are not aligned
47 >>> c.shape
48 (6, 1)
```

Vemos en la columna izquierda cómo podemos construir arreglos, bien directamente, bien disponiendo sus elementos en un orden distinto con el módulo `reshape`. Lo más notable de la diferencia entre el uso de arreglos y matrices es que el operador de multiplicación `*` no corresponde a la multiplicación matricial de arreglos bidimensionales, de ahí el error que aparece en la línea 28. Si observamos la línea 31, vemos que la multiplicación entre arreglos se realiza elemento a elemento, y que la multiplicación matricial se hace con la función `dot`.

Por otra parte, hay que tener cuidado pues no es lo mismo un arreglo unidimensional, que un arreglo bidimensional que tenga una de sus dimensiones igual a uno, como se aprecia en la línea 44. El producto matricial de `c` por `a` no se puede hacer pues `c` es un arreglo bidimensional de orden $6 \times 1$, pero `a` es unidimensional (línea 8).

Sin embargo, cuando se trata de multiplicar por un vector, es decir, por un arreglo unidimensional, la función `dot` permite cierta flexibilidad:

```
>>> d=array((2,3,5)
>>> e=array((5,9))
>>> f=d.reshape(3,1)
>>> dot(b,d)
array([13, 43])
>>> dot(b,f)
array([[13],
       [43]])
>>> dot(e,b)
array([27, 41, 55])
```

Para terminar, si bien comentamos anteriormente que no hay forma directa de calcular el rango de una matriz, dimos una forma sencilla de obtenerlo indirectamente con SymPy, y también se puede lograr con NumPy, usando una función de `linalg` que veremos con más detenimiento en un tema posterior.

```
>>> from numpy import array,linalg
>>> a=array([[2,3,1,-4,2],[1,2,1,-2,3],[2,3,0,2,1],[1,2,0,2,1]])
>>> linalg.lstsq(a,a)[2]
4
```

La función usada es `linalg.lstsq`, y la llamada puede parecer un poco extraña, pues precisa de dos argumentos, el primero de los cuales es el arreglo bidimensional cuyo rango queremos calcular, y el segundo no necesitamos explicarlo aquí (se verá posteriormente en el tema 8); simplemente volvemos a usar el mismo argumento. Esta función proporciona cuatro resultados a través de un arreglo (de forma similar al uso del módulo `LUdecomposition` en la sección 3.4), de los cuales el tercero (índice 2) es precisamente el rango de la matriz introducida.

Dado que Python es un lenguaje modular construido a partir de funciones, podemos aprovecharnos de ello para definir nuestra propia función `rango` que podamos usar en cualquier momento, del mismo modo que tenemos a nuestra disposición el módulo NumPy. La forma de proceder es la misma que hicimos para ejecutar el código de la función `jacobi` al final de la sección 3.4.

Creamos un archivo con el contenido de la función.

```
from numpy import linalg

def rango(a):
        x=linalg.lstsq(a,a)[2]
        return x
```

y lo guardamos con el nombre `mimodulo.py`. Lo más efectivo es guardar este fichero en alguna de las carpetas en las que el intérprete Python busca a la

hora de importar un módulo. Estas carpetas son las definidas en la variable de entorno PYTHONPATH que se puede obtener mediante la función path del módulo sys, y lógicamente depende del sistema operativo que se esté usando. La otra opción es guardar el archivo en la carpeta desde la que ejecutamos el intérprete. Una vez guardado, se procedería así,

```
>>> from numpy import array
>>> a=array([[1,0,1,0],[2,1,3,1],[0,1,1,1],[2,2,4,2]])
>>> from mimodulo import rango
>>> rango(a)
2
```

4 7

## APLICACIÓN: LIGHTS OUT!, PRIMERA PARTE

Vamos a dedicar esta sección al estudio algebraico de un juego electrónico denominado *Lights Out!*, que podría traducirse por *¡Apaga las luces!*, y que consiste en un tablero que consta de $5 \times 5$ casillas (o bombillas), cada una de las cuales puede estar apagada o encendida. Al pulsar sobre cualquier casilla, ésta cambia de estado, conjuntamente con las casillas adyacentes de su línea horizontal y vertical. El juego consiste en, dada una configuración inicial, ir pulsando las casillas necesarias hasta conseguir apagar todas ellas. La figura 4.1 muestra un par de ejemplos del funcionamiento del juego.

Hemos de tener en cuenta dos observaciones importantes de la dinámica de este juego: en primer lugar, al pulsar dos veces sobre la misma casilla el tablero permanece en el mismo estado inicial; y en segundo lugar, es fácil darse cuenta de que el orden en el que se pulsen las casillas no afecta al estado obtenido.

Para analizar matemáticamente el funcionamiento del juego hemos de situarnos en el contexto adecuado, que en este caso corresponde a un e.v. peculiar: $\mathbb{Z}_2^{25}$.

Recordemos que en la definición de e.v. es preciso que exista un cuerpo con el que realizar la operación producto por escalar. Hasta ahora, los cuerpos conocidos por el lector son $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$, pero aquí vamos a necesitar un nuevo conjunto que tiene esta estructura: $\mathbb{Z}_2$, definido como el conjunto de números enteros *módulo* 2.

En general $\mathbb{Z}_p$, el conjunto de números enteros módulo $p$, donde $p \in \mathbb{Z}$, está formado por todos los números enteros en los que se da una cierta relación de equivalencia: diremos que $a$ y $b$ son equivalentes si el resto de la división de ambos entre $p$ es el mismo. Se notará por $a \equiv b \pmod{p}$. Por ejemplo $25 \equiv 35 \pmod{10}$, ó $3 \equiv 1 \pmod{2}$. Es decir, $\mathbb{Z}_p$ es el conjunto cociente definido por esta relación de equivalencia (véase el Apéndice A).

A partir de aquí es fácil observar que $Z_p = \{0, 1, 2, \dots, p-1\}$, es decir, es un conjunto finito. Si además $p$ es un número primo, la operación suma y

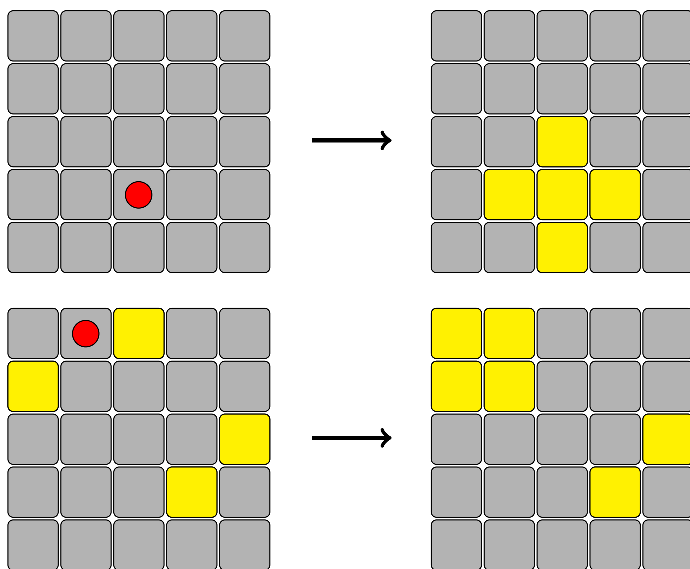


Figura 4.1: Así se juega a *Lights Out!*

producto habituales en $\mathbb{Z}$ dotan a $\mathbb{Z}_p$ de estructura de cuerpo. Por ejemplo, si consideramos $Z_2 = \{0, 1\}$, entonces la operación suma en este conjunto es

$$0 + 0 = 0, \quad 0 + 1 = 1, \quad 1 + 0 = 1, \quad 1 + 1 = 0 \tag{4.12}$$

El lector quizás se pregunte por qué $1+1 = 0$. La respuesta es sencilla, $1+1 = 2$, pero $2 \equiv 0 \pmod 2$, pues el resto de la división de 2 entre 2 es 0; así funciona lo que se denomina como *aritmética modular*. La operación suma definida en (4.12) junto con la operación producto de enteros en este espacio $\mathbb{Z}_2$ satisfacen las propiedades adecuadas para que $\mathbb{Z}_2$ sea un cuerpo.

Si ahora consideramos $\mathbb{Z}_2^n$, el producto cartesiano de $\mathbb{Z}_2$ consigo mismo, $n$ veces, y lo dotamos de la suma y el producto escalar habituales, en el que el cuerpo de escalares corresponde a $\mathbb{Z}_2$, entonces tenemos una estructura de espacio vectorial. Este e.v. tiene la peculiaridad de constar solo de un número finito de elementos (en concreto $2^n$).

Volvamos al juego: el estado de encendido o apagado de cada una de las casillas del tablero se puede modelar en $\mathbb{Z}_2$, asignando un 0 a la casilla apagada y un 1 a la casilla encendida. Así, la suma descrita en (4.12) corresponde al cambio de estado de una casilla. Como tenemos 25 casillas, vamos a establecer el estado de todo el tablero como un vector de $Z_2^{25}$, en el que consideraremos la numeración dada en la figura 4.2.

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 |

Figura 4.2: Numeración de las casillas

Así, por ejemplo, el vector que describe el último tablero de la figura 4.1 es

$$(1,1,0,0,0,1,1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0)$$

De este modo, un vector cualquiera de $\mathbb{Z}_2^{25}$ corresponderá a una configuración determinada de luces encendidas y apagadas del tablero. El objetivo del juego consistirá entonces en, dada una configuración inicial, encontrar qué combinación de pulsaciones nos lleva a la configuración nula.

Ahora bien, atendiendo a las observaciones que hicimos antes, si tenemos una configuración inicial $\mathbf{b}$ obtenida al presionar una serie de casillas en el tablero inicialmente apagado, volviendo a presionar nuevamente esas casillas conseguiremos apagar todo el tablero. Esta sucesión de pulsaciones, lo que se denomina una estrategia, puede ser representada también por otro vector de $\mathbb{Z}_2^{25}$,

$$\mathbf{x} = (x_1, x_2, \dots, x_{24}, x_{25})$$

en el que $x_i$ será 1 si se ha pulsado la casilla $i$ y 0 si no se ha pulsado. Entonces, dado $\mathbf{b} \in \mathbb{Z}_2^{25}$, la estrategia ganadora verifica:

$$\begin{aligned}
b_1 &= x_1 + x_2 + x_6 \\
b_2 &= x_1 + x_2 + x_3 + x_7 \\
b_3 &= x_2 + x_3 + x_4 + x_8 \\
&\vdots \\
b_{24} &= x_{19} + x_{23} + x_{24} + x_{25} \\
b_{25} &= x_{20} + x_{24} + x_{25}
\end{aligned}$$

Es fácil reescribir el sistema anterior en forma matricial, $A\mathbf{x} = \mathbf{b}$, donde la

matriz $A \in \mathcal{M}_{25}(\mathbb{Z}_2)$ tiene la siguiente estructura de bloques

$$A = \begin{pmatrix} B & I & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I & B & I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & B & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I & B & I \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I & B \end{pmatrix} \tag{4.13}$$

donde $\mathbf{0}$ denota la matriz nula de tamaño $5 \times 5$, $I$ es la matriz identidad de tamaño $5 \times 5$ y $B$ es

$$B = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Atendiendo al significado del producto de matrices, podemos observar que $A\mathbf{x}$ corresponde a una combinación lineal de los vectores que conforman las columnas de la matriz $A$, y en la que los escalares son los elementos del vector $\mathbf{x}$. Así pues, $\mathbf{b}$ es una configuración alcanzable si dicho vector es combinación lineal de los vectores columna de $A$. En la sección 8.5 volveremos sobre este juego para averiguar la forma de saber si una configuración inicial puede ser apagada o no de una forma más rápida.

4 8

## EJERCICIOS

### Ejercicios de repaso

E.1 Determinar si los siguientes conjuntos de vectores son linealmente dependientes o independientes. En caso de que sean linealmente dependientes encontrar una combinación lineal no nula entre ellos que proporcione el vector nulo:

(a) $\{(3, 5, 1), (2, 1, 3)\}$ (b) $\{(1, 2, 3), (1, 3, 2), (0, -1, 1)\}$

(c) $\{(1, 0, 1, 0), (2, 1, 3, 1), (0, 1, 1, 1), (2, 2, 4, 2)\}$

E.2 En cada uno de los apartados del ejercicio 1 calcular el rango del conjunto de vectores dado.

E.3 Estudiar si las siguientes familias de vectores son o no l.i.:

(a) $\left\{ \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right\} \subset \mathcal{M}_2(\mathbb{R})$.

(b) $\{x-1, x+1, x^2-1\} \subset \mathbb{P}^2_{\mathbb{R}}[x]$.

(c) $\{\operatorname{sen}^2(\pi x), \cos^2(\pi x), \cos(2\pi x)\} \subset C([0,1])$.

**E.4** Estudiar si los siguientes conjuntos son bases del espacio vectorial dado:

(a) $\{1, x+3, (x+3)^2, (x+3)^3\}$ en $\mathbb{P}^3_{\mathbb{R}}[x]$.

(b) $\left\{\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\right\}$ en $\mathcal{M}_2(\mathbb{R})$.

(c) $\left\{\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}, \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}, \begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}, \begin{pmatrix}-1 & 0\\ 0 & 0\end{pmatrix}\right\}$ en $\mathcal{M}_2(\mathbb{R})$.

**E.5** En $\mathbb{R}^3$ se consideran los siguientes conjuntos:

$$A = \{(1,5,1), (2,1,0)\}$$
$$C = \{(1,5,1), (2,1,0), (1,0,0), (0,0,1)\}$$

Se pide:

(a) Probar que $A$ es l.i. y que $C$ es un sistema generador de $\mathbb{R}^3$.

(b) Encontrar una base $B$ que contenga a $A$ y esté contenida en $C$.

**E.6** Demostrar que el conjunto $S = \{x+1, x-1, x^2-1, x^2+1\}$ es un sistema generador de $\mathbb{P}^2_{\mathbb{R}}[x]$. Encontrar un subconjunto $S_1$ de $S$ que sea una base de $\mathbb{P}^2_{\mathbb{R}}[x]$.

**E.7** Sea $\mathcal{B}$ la base canónica de $\mathbb{R}^3$ y consideremos el conjunto

$$\mathcal{B}' = \{(1,1,1), (0,1,1), (0,0,1)\}$$

(a) Demostrar que $\mathcal{B}'$ es una base de $\mathbb{R}^3$.

(b) Hallar las coordenadas del vector $\mathbf{u} = (3,-2,5)$ respecto de la base $\mathcal{B}'$.

(c) Hallar las ecuaciones del cambio de base de $\mathcal{B}$ a $\mathcal{B}'$.

**E.8** Verificar si los siguientes conjuntos son subespacios vectoriales:

(a) $\{(x_1,\dots,x_n) \in \mathbb{R}^n : x_1 + \cdots + x_n = 0\}$.

(b) $\{(x_1,\dots,x_n) \in \mathbb{R}^n : x_1 + \cdots + x_n = 1\}$.

(c) $\{p(x) \in \mathbb{P}^3_{\mathbb{R}}[x] : (x-1) \text{ divide a } p(x)\}$.

(d) $\operatorname{COM}(B) = \left\{A \in \mathcal{M}_2(\mathbb{R}) : AB = BA \text{ donde } B = \begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}\right\}$.

**E.9** En $\mathbb{R}^3$ consideramos el subespacio generado por los vectores $(2,3,0)$ y $(4,6,1)$. ¿Pertenece el vector $(1,0,0)$ a este subespacio?

**E.10** En $\mathbb{R}^3$ consideramos los siguientes conjuntos:

$$A = \{(1,1,1),(0,1,0)\} \quad \text{y} \quad B = \{(2,3,2),(1,0,1)\}.$$

¿Es $L(A) = L(B)$?

**E.11** En $\mathbb{R}^4$ determinar la dimensión y una base del subespacio vectorial formado por las soluciones de los siguientes sistemas:

$$\text{(a)} \begin{cases} x_1 + x_2 - x_3 - x_4 = 0 \\ x_1 - x_2 + 2x_3 - x_4 = 0 \end{cases} \qquad \text{(b)} \begin{cases} x_1 + x_2 + 2x_3 - x_4 = 0 \\ x_1 + x_2 + 3x_3 + 2x_4 = 0 \end{cases}$$

$$\text{(c)} \begin{cases} 3x_1 + x_2 + 4x_3 - 2x_4 = 0 \\ x_1 + 2x_2 + 3x_3 + x_4 = 0 \\ 2x_1 - x_2 + x_3 - 3x_4 = 0 \\ x_1 + x_2 + 2x_3 = 0 \end{cases}$$

**E.12** Sea $V$ un espacio vectorial real de dimensión 5. Para cada uno de los subespacios siguientes calcular un sistema de ecuaciones implícitas:

(a) $L_1 = L((1,0,1,-1,1),(1,1,1,0,0))$.

(b) $L_2 = L((0,1,2,1,0),(1,1,-1,-2,1),(3,-1,-7,-8,3))$.

## Problemas

**E.13** En un espacio vectorial $V$ se tiene una base $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ y un vector $\mathbf{x}$ cuyas coordenadas con respecto a $\mathcal{B}$ son $(1,-1,2)$. Demostrar que el conjunto

$$\mathcal{S} = \{\mathbf{u}_1 + \mathbf{u_2}, \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3\}$$

es l.i. y completar $\mathcal{S}$ a una base $\mathcal{B}'$ tal que las coordenadas de $\mathbf{x}$ con respecto a $\mathcal{B}'$ sean $(1,1,1)$.

**E.14** Sabiendo que un vector $\mathbf{v} \in \mathbb{R}^2$ tiene coordenadas $(1,\alpha)$ en la base $\mathcal{B}_1 = \{(1,2),(4,-1)\}$ y $(6,\beta)$ en la base $\mathcal{B}_2 = \{(1,1),(1,-1)\}$, determinar las coordenadas de $\mathbf{v}$ en la base canónica.

**E.15** En $\mathbb{P}^3_{\mathbb{R}}[x]$ consideremos el polinomio $p(x) = ax^3 + bx^2 + cx + d$ con $a \neq 0$).

(a) Demostrar que el conjunto $\mathcal{B} = \{p(x), p'(x), p''(x), p'''(x)\}$ es una base de $\mathbb{P}^3_{\mathbb{R}}[x]$.

(b) Si $p(x) = 5x^3 + 3x^2 - 2x + 1$, hallar las ecuaciones del cambio de base de $\mathcal{B}$ a la base canónica de $\mathbb{P}^3_{\mathbb{R}}[x]$, y las coordenadas del vector $q(x) = 15x^3 - 21x^2 - 18x + 37$ en ambas bases.

**E.16** Consideremos los conjuntos de $\mathbb{R}^3$

$$\mathcal{B}_1 = \{(1,1,1),(0,1,1),(0,0,1)\}, \quad \mathcal{B}_2 = \{(0,1,0),(1,1,-1),(-2,0,1)\}$$

(a) Probar que ambos conjuntos son bases de $\mathbb{R}^3$.

(b) Determinar el conjunto de vectores de $\mathbb{R}^3$ que poseen las mismas coordenadas respecto de ambas bases.

***E.17** Consideremos $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ la base canónica de $\mathbb{R}^n$ y los vectores

$$\mathbf{u}_1 = \mathbf{e}_2 - \mathbf{e}_1, \ \mathbf{u}_2 = \mathbf{e}_3 - \mathbf{e}_2, \dots, \mathbf{u}_{n-1} = \mathbf{e}_n - \mathbf{e}_{n-1}, \ \mathbf{u}_n = \mathbf{e}_n$$

(a) Demostrar que los $\mathbf{u}_j$ forman base en $\mathbb{R}^n$.

(b) Encontrar las coordenadas del vector $\mathbf{x} = \sum_{i=1}^n \mathbf{e}_i$ con respecto a la base formada por los $\mathbf{u}_j$.

**E.18** ¿Para qué valores de $\alpha$ y $\beta$ el conjunto $W = \{p(x) \in \mathbb{P}^2_{\mathbb{R}}[x] : p(0) = \alpha, \ p'(0) = \beta\}$ es un subespacio vectorial? Para los valores obtenidos obtener una base y la dimensión de dicho subespacio.

**E.19** Determinar la dimensión del subespacio vectorial formado por las soluciones de los siguientes sistemas, en función del parámetro $m \in \mathbb{R}$:

(a) $\begin{cases} x + m^2y + mz = 0 \\ mx + y + mz = 0 \\ m^2x + my + mz = 0 \end{cases}$ (b) $\begin{cases} (4-m)x + (1-m)y + z = 0 \\ (m-1)x + (2m-1)y + 2z = 0 \\ (5-m)x + my - z = 0 \end{cases}$

**E.20** Calcular la dimensión y una base de los siguientes espacios:

(a) $W = \{p(x) \in \mathbb{P}^2_{\mathbb{R}}[x] : p(0) = 0, \ p(2) = 0\}$.

(b) El subconjunto de $\mathcal{M}_{2\times 3}(\mathbb{R})$ tales que la suma de los elementos de cada columna suman cero.

**E.21** En $\mathbb{R}^4$ se consideran los pares de subespacios que siguen. Hallar la dimensión y una base de $L_1$, $L_2$, $L_1 + L_2$ y $L_1 \cap L_2$. ¿En qué casos se tiene que $\mathbb{R}^4 = L_1 \oplus L_2$?

(a) $L_1 = L((1,1,1,1),(1,-1,1,-1),(1,3,1,3))$;
$L_2 = L((1,2,0,2),(1,2,1,2),(3,1,3,1))$.

(b) $L_1 = L((1,-1,2,1),(0,1,-1,3),(2,0,1,-1))$;
$L_2 = \begin{cases} 2x_1 - x_2 - 3x_3 = 0 \\ x_1 - 2x_2 + 6x_3 - 6x_4 = 0 \end{cases}$

(c) $L_1 = \begin{cases} 3x_1 + 2x_2 + 3x_3 - x_4 = 0 \\ -x_1 + 5x_3 + x_4 = 0 \\ 6x_1 + 4x_2 + 6x_3 - 2x_4 = 0 \end{cases}$

$L_2 = \{ \; x_2 = 0$

**E.22** Hallar la suma y la intersección de cada uno de los siguientes pares de subespacios. Estudiar, cuando proceda, su suma directa:

(a) $V_1 = \mathrm{COM}(A)$ donde $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, $V_2 = \{A \in \mathcal{M}_2(\mathbb{R}) : \mathrm{tr}(A) = 0\}$

Nota: el espacio $\mathrm{COM}(A)$ se definió en el ejercicio 8, mientras que $\mathrm{tr}(A)$ denota la *traza* de un matriz, definida como la suma de sus elementos diagonales.

(b) $V_1$ es el conjunto de matrices simétricas en $\mathcal{M}_2(\mathbb{R})$,
$V_2$ es el conjunto de matrices antisimétricas en $\mathcal{M}_2(\mathbb{R})$.

(c) $V_1 = \{p(x) \in \mathbb{P}^3_{\mathbb{R}}[x] : (x-1) \text{ divide a } p(x)\}$,
$V_2 = \{p(x) \in \mathbb{P}^3_{\mathbb{R}}[x] : (x+1) \text{ divide a } p(x)\}$.

*(d) $V_1$ es el conjunto de funciones $f : \mathbb{R} \to \mathbb{R}$ pares en $\mathcal{C}(\mathbb{R})$,
$V_2$ es el conjunto de funciones $f : \mathbb{R} \to \mathbb{R}$ impares en $\mathcal{C}(\mathbb{R})$

Nota: $\mathcal{C}(\mathbb{R})$ denota el conjunto de funciones continuas en $\mathbb{R}$.

**E.23** Sean $L_1$ y $L_2$ dos subespacios vectoriales de un e.v. $V$ tal que $\dim(V) = 5$ y $\dim(L_1) = \dim(L_2) = 3$. ¿En cuáles de las siguientes situaciones se cumple que para todo vector $\mathbf{u} \in V$, existen vectores $\mathbf{w}_1 \in L_1$ y $\mathbf{w}_2 \in L_2$ tales que $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$?

(a) Siempre, aunque los vectores $\mathbf{w}_1$ y $\mathbf{w}_2$ pueden ser no únicos.

(b) Cuando $\dim(L_1 \cap L_2) = 1$.

(c) Cuando $\dim(L_1 \cap L_2) = 2$.

(d) Cuando $V = L_1 + L_2$, en cuyo caso los vectores $\mathbf{w}_1$ y $\mathbf{w}_2$ son únicos.

### Ejercicios teóricos

**E.24** Definimos en $\mathbb{R}^2$ las operaciones suma y producto por un escalar real del siguiente modo:

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2), \quad \alpha\,(x_1, x_2) = (\alpha^2 x_1, \alpha^2 x_2), \; \alpha \in \mathbb{R}$$

Estudiar si, con estas operaciones, $\mathbb{R}^2$ es un espacio vectorial.

* **E.25** Probar que $\mathbb{C}^2$ es un espacio vectorial real y calcula su dimensión. ¿Es diferente si se considera $\mathbb{C}^2$ como e.v. complejo?

***E.26** Sean los polinomios $p_1(x)$, $p_2(x)$, …, $p_k(x)$ y supongamos que

$$\begin{vmatrix} p_1(a_1) & p_2(a_1) & \cdots & p_k(a_1) \\ p_1(a_2) & p_2(a_2) & \cdots & p_k(a_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(a_k) & p_2(a_k) & \cdots & p_k(a_k) \end{vmatrix} \neq 0$$

para ciertos números reales $a_j$ distintos entre sí. Demostrar que el conjunto de polinomios dado es linealmente independiente en $\mathbb{P}_{\mathbb{R}}[x]$.

***E.27** Sea $A \in \mathcal{M}_n(\mathbb{K})$ una matriz no singular y $\{\mathbf{v}_1, \dots \mathbf{v}_n\}$ un conjunto de vectores de $\mathbb{K}^n$ linealmente independientes. Probar que el conjunto $\{A\mathbf{v}_1, \dots A\mathbf{v}_n\}$ es linealmente independiente. ¿Es cierto el resultado si $A$ es singular?

**E.28** Probar el Teorema 4.11.

**E.29** Sea $V$ un espacio vectorial. Probar que si la dimensión de la suma de dos subespacios de $V$ es una unidad mayor que la dimensión de su intersección entonces la suma coincide con uno de ellos y la intersección con el otro.

**E.30** Si $L_1$ y $L_2$ son subespacios de un e.v. $V$ tales que $V = L_1 \oplus L_2$, probar que si $\mathcal{B}_1$ y $\mathcal{B}_2$ son bases de $L_1$ y $L_2$, respectivamente, entonces $\mathcal{B}_1 \cup \mathcal{B}_2$ es base de $V$.

### Ejercicios adicionales

**E.31** Considérese un juego idéntico a *Lights out!* pero con un tablero formado solo por cuatro casillas, como el siguiente:

| | |
|---|---|
| 1 | 2 |
| 3 | 4 |

Al pulsar sobre una casilla, cambia el estado de esa casilla, así como sus adyacentes (en horizontal y vertical). Plantear el sistema de ecuaciones correspondiente a este tablero y averiguar si existe solución para la siguiente configuración inicial:

resolviendo el sistema mediante el método de Gauss, usando la aritmética modular descrita en (4.12).

# 5 Aplicaciones lineales

Las funciones constituyen sin duda alguna el objeto matemático de mayor interés y uso en cualquier disciplina en la que las matemáticas tengan un mínimo de presencia. En este tema iniciaremos el estudio de las funciones relacionadas con los espacios vectoriales, más conocidas como *aplicaciones lineales*.

## 5.1 APLICACIONES LINEALES ENTRE ESPACIOS VECTORIALES

Básicamente podemos decir que una aplicación lineal es aquella que respeta la estructura de espacio vectorial:[1]

**Definición 5.1**

Sean $V$ y $W$ dos espacios vectoriales sobre un mismo cuerpo $\mathbb{K}$. Una aplicación $f : V \to W$ se dice *lineal* si verifica:

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in V$$
$$f(\alpha\mathbf{x}) = \alpha f(\mathbf{x}) \quad \forall \mathbf{x}, \in V, \quad \forall \alpha \in \mathbb{K}$$

De forma equivalente, $f$ es lineal si y sólo si $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$, $\forall \mathbf{x}, \mathbf{y} \in V$, $\forall \alpha, \beta \in \mathbb{K}$.

Como puede verse en la definición, una aplicación lineal no es más que una función definida entre espacios vectoriales que respecta la linealidad, es decir, la imagen de una combinación lineal es la combinación lineal de las imágenes. A pesar de su simplicidad, su importancia es enorme en multitud de disciplinas. Veamos algunos ejemplos:

[1] Las transformaciones lineales ya aparecen en la ya citada obra de Grassmann que pasó casi desapercibida en su época, aunque el concepto de *linealidad* aparece frecuentemente con anterioridad en numerosas aplicaciones.

**Ejemplo 5.1**

(I) La aplicación *nula*, definida por

$$\begin{array}{rccc} 0_V : & V & \longrightarrow & V \\ & \mathbf{v} & \longrightarrow & \mathbf{0} \end{array}$$

es lineal.

(II) La aplicación *identidad* dada por

$$\begin{array}{rccc} I_V : & V & \longrightarrow & V \\ & \mathbf{v} & \longrightarrow & \mathbf{v} \end{array}$$

es lineal.

(III) La simetría plana de eje la recta $y = x$, cuya ecuación es

$$\begin{array}{rccc} \sigma : & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (x, y) & \longrightarrow & (y, x) \end{array}$$

es una aplicación lineal.

(IV) La aplicación derivada es lineal:

$$\begin{array}{rccc} d : & \mathbb{P}_{\mathbb{R}}[x] & \longrightarrow & \mathbb{P}_{\mathbb{R}}[x] \\ & p(x) & \longrightarrow & p'(x) \end{array}$$

Algunos ejemplos de aplicaciones no lineales:

$$\begin{array}{rccc} f_1 : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \longrightarrow & x^2 \end{array}$$

$$\begin{array}{rccc} f_2 : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \longrightarrow & \operatorname{sen} x \end{array}$$

Notaremos por $\mathcal{L}(V, W)$ al conjunto de aplicaciones lineales de $V$ en $W$ (también llamadas *homomorfismos*). Si $V = W$ se dice que $f$ es un *endomorfismo*, y al conjunto de tales aplicaciones se le denota por $\mathcal{L}(V)$.

Se dice que $f : V \longrightarrow W$ es un *isomorfismo* si es lineal y biyectiva. También

se dice que $f$ es un *automorfismo* si es un endomorfismo biyectivo.

### 5 1 1 Propiedades

De la definición de aplicación lineal se deduce fácilmente que la imagen de una combinación lineal de vectores es igual a la combinación lineal de las imágenes, es decir,

$$f\left(\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\right) = \sum_{i=1}^{n} \alpha_i f(\mathbf{u}_i) \tag{5.1}$$

Algunas consecuencias inmediatas son:

**Proposición 5.1**

Si $f : V \longrightarrow W$ es lineal, entonces,

(I) $f(\mathbf{0}) = \mathbf{0}$.

(II) $f(-\mathbf{u}) = -f(\mathbf{u})$.

(III) Si $L$ es un subespacio vectorial, $f(L) = \{f(\mathbf{v}) : \mathbf{v} \in L\}$ también es un subespacio vectorial.

La demostración de este resultado es elemental y se propone al lector como ejercicio.

**Proposición 5.2**

Sea $L$ un subespacio de dimensión $k$ y $f$ una aplicación lineal, entonces $f(L)$ es un subespacio de dimensión menor o igual que $k$.

*Demostración:*
Sea $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ una base de $L$. Si $\mathbf{y} \in f(L)$ entonces existe $\mathbf{x} \in L$ tal que $f(\mathbf{x}) = \mathbf{y}$. Como $\mathbf{x} \in L$, podemos escribir que $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i$. Usando (5.1),

$$\mathbf{y} = f\left(\sum_{i=1}^{n} \alpha_i \mathbf{e}_i\right) = \sum_{i=1}^{n} \alpha_i f(\mathbf{e}_i)$$

esto es, $\mathbf{y}$ es combinación lineal de los vectores $f(\mathbf{e}_1), \dots, f(\mathbf{e}_k)$. Por tanto estos vectores forman un sistema generador de $f(L)$. Del Corolario 4.5 se sigue el resultado.

■

El siguiente resultado es esencial para la determinación de una aplicación lineal, y además pone de manifiesto la importancia de las bases en los e.v.

**Teorema 5.1**

Sea $\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ una base de $V$. Sean $\mathbf{w}_1, \dots, \mathbf{w}_n$ $n$ vectores cualesquiera de $W$. Entonces existe una única aplicación lineal $f : V \longrightarrow W$ tal que $f(\mathbf{e}_i) = \mathbf{w}_i$, $1 \leq i \leq n$.

*Demostración:*
Sea $\mathbf{v} \in V$. Para definir $f(\mathbf{v})$ procedemos como sigue. Puesto que $\mathcal{B}$ es una base de $V$ existen unos únicos $\alpha_i$ tales que $\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i$. En tal situación definimos

$$f(\mathbf{v}) = \sum_{i=1}^{n} \alpha_i \mathbf{w}_i$$

Es fácil comprobar que $f$ es lineal y puesto que las coordenadas respecto de una base son únicas, $f$ está definida de manera única.

■

Como consecuencia, podemos afirmar que para determinar completamente una aplicación lineal basta conocer las imágenes de los elementos de una base del espacio de partida. Con esta simple información (la imagen de unos cuantos vectores) podemos obtener la imagen de cualquier otro vector del espacio.

**Ejemplo 5.2**

Supongamos que conocemos que una aplicación lineal satisface

$$f(1,0) = (2,1,3), \quad f(0,1) = (4,-1,0)$$

Vamos a determinar la aplicación. En primer lugar se trata de una aplicación $f : \mathbb{R}^2 \to \mathbb{R}^3$, y puesto que sabemos los valores de la aplicación sobre una base de $\mathbb{R}^2$, esto es, conocemos la imagen de $\{(1,0),(0,1)\}$, podemos calcular la imagen de cualquier vector $(x_1, x_2)$.

En efecto, usando coordenadas se tiene que

$$(x_1, x_2) = x_1(1,0) + x_2(0,1)$$

luego,

$$\begin{aligned} f(x_1,x_2) &= f(x_1(1,0) + x_2(0,1)) \\ &= x_1 f(1,0) + x_2 f(0,1) = x_1(2,1,3) + x_2(4,-1,0) \\ &= (2x_1 + 4x_2, x_1 - x_2, 3x_1) \end{aligned}$$

5 2

## MATRIZ DE UNA APLICACIÓN LINEAL.

Ya hemos visto que para determinar completamente una aplicación lineal nos basta con conocer las imágenes de una base del espacio de partida. A continuación vamos a usar esta idea para describir con mayor eficacia una aplicación lineal.

Consideremos $V$ y $W$ dos e.v., $f : V \to W$ una aplicación lineal, y sean

$$\mathcal{B} = \{\mathbf{u}_1, \dots \mathbf{u}_n\} \quad \text{y} \quad \mathcal{B}' = \{\mathbf{v}_1, \dots \mathbf{v}_m\}$$

bases de $V$ y $W$, respectivamente.

Dado un vector $\mathbf{x} \in V$ podemos escribir $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{u}_j$, usando sus coordenadas. Entonces

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j f(\mathbf{u}_j) = \mathbf{y} \tag{5.2}$$

Obviamente $\mathbf{y} \in W$, de modo que, haciendo uso de la base de $W$, $\mathbf{y} = \sum_{i=1}^{m} y_i \mathbf{v}_i$. El objetivo que perseguimos es relacionar las coordenadas del vector de partida $\mathbf{x}$ con las coordenadas del vector de llegada $\mathbf{y}$.

Nótese que cada vector $f(\mathbf{u}_j) \in W$, por lo tanto, puesto que tenemos una base de $W$, es posible escribir

$$f(\mathbf{u}_j) = \sum_{i=1}^{m} a_{ij} \mathbf{v}_i, \quad j = 1, \dots, n$$

es decir, $a_{ij}$, $1 \leq i \leq m$ son las coordenadas de la imagen de $f(\mathbf{u}_j)$ en la base de $W$.

Usando esta última observación en (5.2) se tiene que

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j \left( \sum_{i=1}^{m} a_{ij} \mathbf{v}_i \right) = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} x_j a_{ij} \right) \mathbf{v}_i = \sum_{i=1}^{m} y_i \mathbf{v}_i$$

Como las coordenadas son únicas, vemos que

$$y_i = \sum_{j=1}^{n} x_j a_{ij}$$

Si ahora consideramos la matriz $A = (a_{ij})$ entonces la igualdad anterior no es más que el producto de la matriz $A$ por las coordenadas del vector $\mathbf{x}$. Es decir, $f(\mathbf{x}) = A\mathbf{x}$.

A la matriz $A$ se le denomina matriz de $f$ en las bases $\mathcal{B}$ y $\mathcal{B}'$, y se denota por $M_{\mathcal{B}}^{\mathcal{B}'}(f)$. Es importante observar cómo está construida esta matriz: **sus columnas están formadas por las coordenadas, respecto de la base $\mathcal{B}'$,**

**de las imágenes de los vectores de la base** $\mathcal{B}$. Como las dimensiones de $V$ y $W$ son $n$ y $m$, respectivamente, entonces $A \in \mathcal{M}_{m\times n}(\mathbb{K})$.

En conclusión, toda aplicación lineal tiene asociada una matriz respecto de unas bases predeterminadas (nótese que la matriz cambiará si cambiamos de base). En adelante, fijadas las bases, identificaremos la aplicación lineal con la matriz correspondiente, escribiendo $f(\mathbf{x}) = A\mathbf{x}$.

Recíprocamente, sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ y fijemos unas bases $\mathcal{B}$ y $\mathcal{B}'$ en $V$ y $W$, respectivamente. Podemos asumir que las columnas de $A$ son las coordenadas, respecto de la base $\mathcal{B}'$, de las imágenes de los vectores de la base $\mathcal{B}$; de este modo, (véase el Teorema 5.1) queda determinada una única aplicación lineal $f : V \to W$, cuya matriz $M_{\mathcal{B}}^{\mathcal{B}'}(f) = A$.

**Nota 5.1**

Si los espacios $V$ y $W$ coinciden y consideramos la misma base en ambos (que es lo más natural), la matriz de la aplicación correspondiente se notará por $M_{\mathcal{B}}(f)$.

**Ejemplo 5.3 (continuación del ejemplo 5.2)**

Construyamos la matriz de la aplicación dada. Puesto que conocemos las imágenes de la base canónica de $\mathbb{R}^2$

$$f(1,0) = (2,1,3), \qquad f(0,1) = (4,-1,0)$$

directamente podemos escribir la matriz respecto de las bases canónicas de $\mathbb{R}^2$ y $\mathbb{R}^3$, respectivamente, que es

$$\begin{pmatrix} 2 & 4 \\ 1 & -1 \\ 3 & 0 \end{pmatrix}$$

Por tanto, para calcular $f(x_1, x_2)$ no tenemos más que multiplicar por la matriz,

$$f(x_1,x_2) = \begin{pmatrix} 2 & 4 \\ 1 & -1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + 4x_2 \\ x_1 - x_2 \\ 3x_1 \end{pmatrix}$$

Recuérdese que trabajamos con los vectores por columnas (véase la nota 4.6 en la pág. 158).

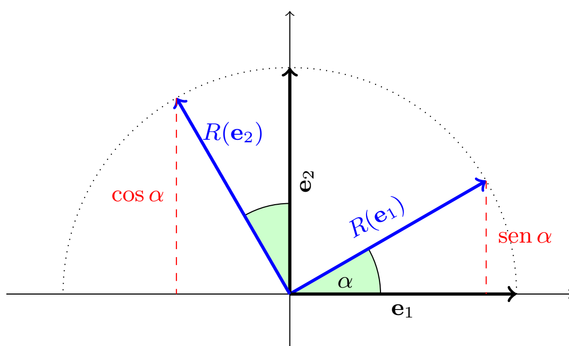



Figura 5.1: Rotación de ángulo $\alpha$

**Ejemplo 5.4**

Matriz de una rotación de ángulo $\alpha$ en $\mathbb{R}^2$, respecto de la base canónica.[2]
Sea $\mathcal{B}_c = \{\mathbf{e}_1, \mathbf{e}_2\}$. Atendiendo a la figura 5.1, se tiene

$$R(\mathbf{e}_1) = \cos(\alpha)\mathbf{e}_1 + \operatorname{sen}(\alpha)\mathbf{e}_2$$
$$R(\mathbf{e}_2) = -\operatorname{sen}(\alpha)\mathbf{e}_1 + \cos(\alpha)\mathbf{e}_2$$

de modo que la matriz de esta rotación en la base canónica de $\mathbb{R}^2$ es

$$M_{\mathcal{B}_c}(R) = \begin{pmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{pmatrix}$$

**Ejemplo 5.5**

Matriz de una simetría respecto del plano $x_3 = 0$, en la base canónica de $\mathbb{R}^3$.
Es fácil comprobar que si $\mathbf{e}_1$, $\mathbf{e}_2$ y $\mathbf{e}_3$ son los vectores de la base canónica de $\mathbb{R}^3$, entonces

$$\sigma(\mathbf{e}_1) = \mathbf{e}_1, \quad \sigma(\mathbf{e}_2) = \mathbf{e}_2, \quad \sigma(\mathbf{e}_3) = -\mathbf{e}_3$$

[2]Hacemos hincapié en el hecho de que la matriz depende de dos bases, la del espacio de partida y la del de llegada, pero en este caso, al tratarse del mismo espacio usamos la misma base.

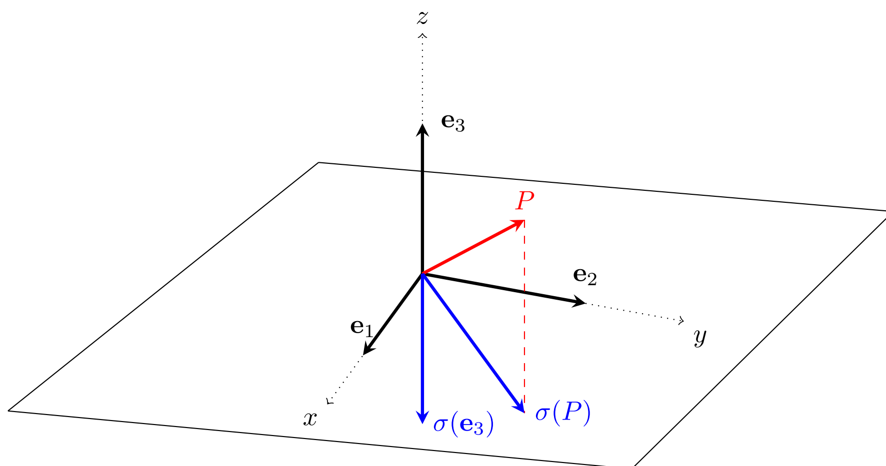



Figura 5.2: Simetría respecto del plano $XY$

(véase la figura 5.2), luego

$$M_{\mathcal{B}_c}(\sigma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Si ahora queremos calcular la imagen por esta simetría de, por ejemplo, el vector $\mathbf{v} = (1, 2, 1)$ bastará multiplicar por la matriz anterior,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

**Ejemplo 5.6**

(i) Matriz de la aplicación lineal

$$\begin{array}{rccc} f: & \mathbb{P}_{\mathbb{R}}^{n}[x] & \longrightarrow & \mathbb{P}_{\mathbb{R}}^{n-1}[x] \\ & p(x) & \longrightarrow & p'(x) \end{array}$$

respecto de las bases $\mathcal{B} = \{1, x, \ldots, x^n\}$ y $\mathcal{B}' = \{1, x, \ldots, x^{n-1}\}$.
Comenzamos calculando las imágenes de los elementos de $\mathcal{B}$ (la base del espacio de partida):

$$f(1) = 0, \quad f(x) = 1, \quad f(x^2) = 2x, \quad \ldots, \quad f(x^n) = nx^{n-1}$$

y a continuación, escribimos estos vectores en coordenadas respecto de $\mathcal{B}'$ (la base del espacio de llegada). Por tanto,

$$M_{\mathcal{B}}^{\mathcal{B}'}(f) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n \end{pmatrix} \in \mathcal{M}_{n\times(n+1)}(\mathbb{R})$$

(ii) Matriz de $g : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por

$$g(x_1, x_2, x_3) = (x_1 + x_2 + x_3, x_1 + x_2 - x_3)$$

respecto de la base canónica de $\mathbb{R}^3$ y la base $\mathcal{B}' = \{(1,0), (1,1)\}$ de $\mathbb{R}^2$.
Calculamos la imagen de los vectores de la base canónica de $\mathbb{R}^3$ y los escribimos como combinación lineal de los vectores de la base de $\mathbb{R}^2$ dada:

$$g(1,0,0) = (1,1) = a_{11}(1,0) + a_{21}(1,1) \Rightarrow a_{11} = 0,\ a_{21} = 1$$
$$g(0,1,0) = (1,1) = a_{12}(1,0) + a_{22}(1,1) \Rightarrow a_{12} = 0,\ a_{22} = 1$$
$$g(0,0,1) = (1,-1) = a_{13}(1,0) + a_{23}(1,1) \Rightarrow a_{13} = 2,\ a_{23} = -1$$

La matriz es por tanto,

$$M_{\mathcal{B}_c}^{\mathcal{B}'}(g) = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 1 & -1 \end{pmatrix}$$

**Nota 5.2**

Obsérvese que el papel que juega la base del espacio de partida en la matriz de una aplicación es completamente diferente al que juega la base del espacio de llegada: la primera es usada para obtener las imágenes a través de la aplicacion de cada uno de sus vectores; mientras que la segunda se emplea para escribir coordenadas respecto de esta base en el espacio de llegada.

5 3

## OPERACIONES ENTRE APLICACIONES LINEALES

En el conjunto de aplicaciones lineales entre dos espacios vectoriales $\mathcal{L}(V,W)$ podemos definir la suma de aplicaciones y el producto con un escalar del siguiente modo:

$$\begin{array}{rccl} +: & \mathcal{L}(V,W)\times\mathcal{L}(V,W) & \longrightarrow & \mathcal{L}(V,W) \\ & (f,g) & \longrightarrow & f+g \quad \text{donde } (f+g)(\mathbf{v}) = f(\mathbf{v})+g(\mathbf{v}) \end{array}$$

$$\begin{array}{rccl} \cdot: & \mathbb{K}\times\mathcal{L}(V,W) & \longrightarrow & \mathcal{L}(V,W) \\ & (\alpha,f) & \longrightarrow & \alpha\cdot f \quad \text{definida por } (\alpha\cdot f)(\mathbf{v}) = \alpha f(\mathbf{v}) \end{array}$$

**Nota 5.3**

Se puede probar que, con las operaciones anteriores, el espacio $\mathcal{L}(V,W)$ es un espacio vectorial de dimensión $(\dim V)(\dim W)$.

El siguiente resultado muestra que las operaciones entre aplicaciones lineales pueden ser vistas como operaciones entre sus matrices asociadas, respecto de las correspondientes bases.

**Proposición 5.3**

Si $A$ y $B$ son las matrices de las aplicaciones $f$ y $g$ en las bases $\mathcal{B}$ y $\mathcal{B}'$, respectivamente, entonces

(i) $A+B$ es la matriz de la aplicación $f+g$ en las citadas bases.

(ii) $\alpha A$ es la matriz de la aplicación $\alpha\cdot f$ en las mismas bases.

La demostración es elemental y se propone al lector.

Además de las operaciones suma y producto por un escalar, es importante prestar atención a la composición de aplicaciones lineales.

**Proposición 5.4**

Si $f \in \mathcal{L}(V, W)$, $g \in \mathcal{L}(W, X)$, $A = M_{\mathcal{B}_V}^{\mathcal{B}_W}(f)$ y $B = M_{\mathcal{B}_W}^{\mathcal{B}_X}(g)$ entonces

$$BA = M_{\mathcal{B}_V}^{\mathcal{B}_X}(g \circ f).$$

*Demostración:*
Sean $A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ y $B = (b_{ki})_{\substack{1\le k\le p\\ 1\le i\le m}}$ las matrices de las aplicaciones dadas en las bases correspondientes. Esto es, si

$$\mathcal{B}_V = \{\mathbf{e}_1, \dots \mathbf{e}_n\}, \quad \mathcal{B}_W = \{\mathbf{e}'_1, \dots \mathbf{e}'_m\}, \quad \mathcal{B}_X = \{\mathbf{e}''_1, \dots \mathbf{e}''_p\}$$

se tiene que

$$f(\mathbf{e}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{e}'_i, \qquad g(\mathbf{e}'_i) = \sum_{k=1}^{p} b_{ki}\mathbf{e}''_k$$

Entonces,

$$\begin{aligned} (g \circ f)(\mathbf{e}_j) = g(f(\mathbf{e}_j)) &= g\left(\sum_{i=1}^{n} a_{ij}\mathbf{e}'_i\right) = \sum_{i=1}^{n} a_{ij} g(\mathbf{e}'_i) \\ &= \sum_{i=1}^{n} a_{ij} \sum_{k=1}^{p} b_{ki}\mathbf{e}''_k = \sum_{k=1}^{p}\left(\sum_{i=1}^{n} b_{ki}a_{ij}\right)\mathbf{e}''_k \end{aligned}$$

Para terminar basta observar que el paréntesis del último sumando corresponde al elemento $kj$ de la matriz $BA$.

■

El resultado anterior nos dice que la composición de aplicaciones lineales se traduce en la multiplicación de sus matrices asociadas, en el orden en el que escribimos tal composición (que no es el orden en el que actúan). Este hecho se extiende a la aplicación inversa.

**Teorema 5.2**

Sea $f$ una aplicación lineal de $V$ en $W$. Entonces,

(I) La inversa de una aplicación lineal es lineal y $(f^{-1})^{-1} = f$.

(II) $f$ es invertible si y solo si es biyectiva.

(III) Si $f : V \to W$ es biyectiva y tiene a $A \in \mathcal{M}_n(\mathbb{K})$ como matriz asociada respecto de ciertas bases, entonces $f^{-1}$ tiene como matriz asociada respecto de las mismas bases a $A^{-1}$.

*Demostración:*

(I) De la definición de aplicación inversa (véase la sección A.2) es inmediato que $(f^{-1})^{-1} = f$. Veamos además que la inversa es lineal; para ello será necesario probar que

$$f^{-1}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f^{-1}(\mathbf{x}) + \beta f^{-1}(\mathbf{y}) \qquad (5.3)$$

Pongamos que $f^{-1}(\mathbf{x}) = \mathbf{x}'$ y $f^{-1}(\mathbf{y}) = \mathbf{y}'$; es decir $f(\mathbf{x}') = \mathbf{x}$ y $f(\mathbf{y}') = \mathbf{y}$. Como $f$ es lineal,

$$f(\alpha\mathbf{x}' + \beta\mathbf{y}') = \alpha f(\mathbf{x}') + \beta f(\mathbf{y}') = \alpha\mathbf{x} + \beta\mathbf{y}$$

y por definición de inversa, $f^{-1}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{x}' + \beta\mathbf{y}' = \alpha f^{-1}(\mathbf{x}) + \beta f^{-1}(\mathbf{y})$, que es precisamente (5.3).

(II) La suficiencia (es decir, que el hecho de que la aplicación sea biyectiva implica que existe inversa) es inmediata, pues si una aplicación es inyectiva, entonces existe inversa, y además como es sobreyectiva, entonces la inversa está definida para todo vector de $W$.

Para la necesidad (esto es, si la aplicación es invertible, entonces es biyectiva), veamos que $f$ es inyectiva y sobreyectiva. Sean $\mathbf{x}$ e $\mathbf{y}$ en $V$ tales que $f(\mathbf{x}) = f(\mathbf{y})$. En tal caso, $f(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Por otra parte, como $f^{-1}$ es lineal $f^{-1}(\mathbf{0}) = \mathbf{0}$ (según (I), Proposición 5.1), pero además $f^{-1}(\mathbf{0}) = \mathbf{x} - \mathbf{y}$, luego $\mathbf{x} = \mathbf{y}$; es decir, $f$ es inyectiva.

Para la sobreyectividad, sea $\mathbf{y} \in W$. Como existe inversa, $f^{-1}(\mathbf{y}) = \mathbf{x}$, para algún vector $\mathbf{x} \in V$, es decir, $f(\mathbf{x}) = \mathbf{y}$. La arbitrariedad de $\mathbf{y}$ nos da la sobreyectividad.

(III) En primer lugar tengamos en cuenta una consecuencia inmediata de un resultado que probaremos más adelante (Teorema 5.3), del cual se deduce que si $f : V \to W$ es biyectiva, entonces $\dim(V) = \dim(W)$. Teniendo esto presente, consideremos $f : V \to W$ biyectiva y una matriz asociada

a $f$ respecto de ciertas bases en estos espacios que llamaremos $A$. Por el comentario anterior, $A$ debe ser una matriz cuadrada.

Por otra parte, como acabamos de probar, al ser $f$ biyectiva existe inversa y además es lineal, $f^{-1}: W \to V$, por lo que podemos considerar su matriz asociada, $B$, (también cuadrada) respecto de las mismas bases de $W$ y $V$, respectivamente.

Consideremos ahora la composición $f^{-1} \circ f$; esta aplicación resulta la identidad de $V$ en $V$, y al ser lineal (composición de aplicaciones lineales) su matriz asociada respecto de la misma base de $V$ será la matriz identidad, $I$.

Finalmente, usando la Proposición 5.4, la matriz $BA$ corresponderá a la composición de $f$ con $f^{-1}$, de modo que $BA = I$. Dado que $A$ y $B$ son cuadradas y su producto es la identidad, se tiene que $B = A^{-1}$ (véase la Proposición 2.5).

■

**Ejemplo 5.7**

Consideremos las siguientes aplicación lineales:

$f: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ definida por $f(x_1, x_2, x_3) = (x_1 + 2x_2 - x_3, -x_1 + x_2 + 2x_3)$

$g: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por $g(x_1, x_2, x_3) = (3x_1 - x_2, x_1 - x_3)$

$h: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ definida por $h(x_1, x_2) = (2x_1 + x_2, -x_1 - x_2)$

y calculemos $(f+g)$, $(3 \cdot f)$ y $h^{-1}$. Es inmediato comprobar que

$$\begin{aligned}
(f+g)(x_1, x_2, x_3) = &\quad (x_1 + 2x_2 - x_3, -x_1 + x_2 + 2x_3) + (3x_1 - x_2, x_1 - x_3) \\
= &\quad (4x_1 + x_2 - x_3, x_2 + x_3) \\
(3 \cdot f)(x_1, x_2, x_3) = &\quad 3 \cdot (x_1 + 2x_2 - x_3, -x_1 + x_2 + 2x_3) \\
= &\quad (3x_1 + 6x_2 - 3x_3, -3x_1 + 3x_2 + 6x_3)
\end{aligned}$$

Para el cálculo de $h^{-1}$ escribimos $h(x_1, x_2) = (y_1, y_2)$ y despejamos $x_1$ y $x_2$:

$$\left.\begin{aligned} 2x_1 + x_2 &= y_1 \\ -x_1 - x_2 &= y_2 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= y_1 + y_2 \\ x_2 &= -y_1 - 2y_2 \end{aligned}$$

de modo que $h^{-1}(y_1, y_2) = (y_1 + y_2, -y_1 - 2y_2)$.

Veamos ahora cómo obtener estas aplicaciones a través del cálculo con matrices. Consideremos las bases canónicas en $\mathbb{R}^3$ y $\mathbb{R}^2$, y las matrices respecto de estas bases de $f$, $g$ y $h$:

$$M(f) = \begin{pmatrix} 1 & 2 & -1 \\ -1 & 1 & 2 \end{pmatrix}, \quad M(g) = \begin{pmatrix} 3 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad M(h) = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}$$

Nótese que estas matrices se pueden calcular tal y como hemos comentado en la sección 5.2. No obstante, es interesante observar que cuando tenemos expresiones analíticas de las aplicaciones y queremos obtener la matriz respecto de las bases canónicas, ésta se obtiene escribiendo por filas los coeficientes de cada una de las componentes de la expresión analítica.

Ahora es sencillo ver que

$$M(f+g) = \begin{pmatrix} 1 & 2 & -1 \\ -1 & 1 & 2 \end{pmatrix} + \begin{pmatrix} 3 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 4 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$M(3\cdot f) = 3\begin{pmatrix} 1 & 2 & -1 \\ -1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & -3 \\ -3 & 3 & 6 \end{pmatrix}$$

$$M(h^{-1}) = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}$$

5 4

## CAMBIO DE BASE EN UNA APLICACIÓN LINEAL

Hemos visto que la matriz de una aplicación lineal depende de las bases que hayamos elegido en los espacios de partida y de llegada. Ahora veremos qué ocurre cuando realizamos un cambio de bases en los espacios.

Supongamos que tenemos una aplicación lineal $f : V \to W$ y las bases en $V$ y $W$ siguientes:

$$\mathcal{B}_V = \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \qquad \mathcal{B}'_V = \{\mathbf{u}'_1, \dots, \mathbf{u}'_n\}$$
$$\mathcal{B}_W = \{\mathbf{v}_1, \dots, \mathbf{v}_m\} \qquad \mathcal{B}'_W = \{\mathbf{v}'_1, \dots, \mathbf{v}'_m\}$$

Denotemos por $A = M_{\mathcal{B}_V}^{\mathcal{B}_W}(f)$ y $A' = M_{\mathcal{B}'_V}^{\mathcal{B}'_W}(f)$. Nuestro interés está en encontrar la relación entre las matrices $A$ y $A'$.

Para ello usaremos la relación entre las bases de cada uno de los espacios: en concreto, si $C = M_{\mathcal{B}'_V}^{\mathcal{B}_V}$, es la matriz de cambio de base en el espacio $V$ de la base $\mathcal{B}'_V$ a $\mathcal{B}_V$, entonces sabemos que, si $\mathbf{x}$ son las coordenadas en la base $\mathcal{B}_V$ y $\mathbf{x}'$ son las coordenadas en la base $\mathcal{B}'_V$ se tiene que $\mathbf{x} = C\mathbf{x}'$.

Del mismo modo, si consideramos $D = M_{\mathcal{B}_W}^{\mathcal{B}'_W}$ la matriz de cambio de base de la base $\mathcal{B}_W$ a $\mathcal{B}'_W$ en el espacio $W$, entonces $\mathbf{y}' = D\mathbf{y}$, donde $\mathbf{y}$ denota las coordenadas respecto de $\mathcal{B}_W$ e $\mathbf{y}'$ las coordenadas respecto de $\mathcal{B}'_W$. Nótese que como las matrices de cambio de base son invertibles (Teorema 4.6), también podemos escribir $\mathbf{y} = D^{-1}\mathbf{y}'$.

Con estas notaciones se tendrá que $\mathbf{y} = A\mathbf{x}$ e $\mathbf{y}' = A'\mathbf{x}'$.

Ahora es sencillo establecer la relación entre $A$ y $A'$:

$$\mathbf{y} = A\mathbf{x} \Rightarrow D^{-1}\mathbf{y}' = A(C\mathbf{x}') \Rightarrow \mathbf{y}' = DAC\mathbf{x}'$$

luego $A' = DAC$.

El siguiente diagrama puede ayudar a esclarecer mejor la situación:

$$\begin{array}{ccc} f : V & \longrightarrow & W \\ \mathcal{B}_V & \xrightarrow{\ A\ } & \mathcal{B}_W \\ C \uparrow & & \downarrow D \\ \mathcal{B}'_V & \xrightarrow[\ A'\ ]{} & \mathcal{B}'_W \end{array}$$

El camino que enlaza la matriz $A'$ puede ser recorrido en la dirección que usa la matriz $A$. Es decir, $A'$ es equivalente a hacer actuar primero $C$, luego $A$ y finalmente $D$. Como la composición se multiplica en orden inverso al orden de actuación obtenemos la igualdad entre matrices anterior.

**Ejemplo 5.8**

(i) Sea $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ la aplicación lineal dada por $f(x_1, x_2, x_3) = (x_1 + x_2, x_3)$ y consideremos las bases

$$\mathcal{B} = \{(1,1,1), (1,1,0), (1,0,0)\}, \quad \mathcal{B}' = \{(1,1), (1,-1)\},$$

de $\mathbb{R}^3$ y $\mathbb{R}^2$, respectivamente. Calculemos $M_{\mathcal{B}}^{\mathcal{B}'}(f)$.

Para encontrar dicha matriz podemos proceder de dos maneras. Mediante la definición de la matriz de una aplicación respecto de unas bases, debemos calcular la imagen de los elementos de $\mathcal{B}$ y escribirlos como com-

binación lineal de los elementos de $\mathcal{B}'$. Así,

$$f(1,1,1) = (2,1) = a_{11}(1,1) + a_{21}(1,-1) \Rightarrow \begin{cases} 2 = a_{11} + a_{21} \\ 1 = a_{11} - a_{21} \end{cases}$$

$$f(1,1,0) = (2,0) = a_{12}(1,1) + a_{22}(1,-1) \Rightarrow \begin{cases} 2 = a_{12} + a_{22} \\ 0 = a_{12} - a_{22} \end{cases}$$

$$f(1,0,0) = (1,0) = a_{13}(1,1) + a_{23}(1,-1) \Rightarrow \begin{cases} 1 = a_{13} + a_{23} \\ 0 = a_{13} - a_{23} \end{cases}$$

Resolviendo los sistemas obtenemos

$$a_{11} = \tfrac{3}{2},\ a_{21} = \tfrac{1}{2},\ a_{12} = 1,\ a_{22} = 1,\ a_{13} = \tfrac{1}{2},\ a_{23} = \tfrac{1}{2},$$

luego,

$$M_{\mathcal{B}}^{\mathcal{B}'}(f) = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \end{pmatrix}$$

Otra forma de calcular esta matriz con un menor número de operaciones consiste en calcular la matriz de la aplicación respecto de las bases canónicas de ambos espacios (que es inmediata de obtener, como comentamos en el ejemplo 5.7) y efectuar el correspondiente cambio de base. Así, si denotamos por $\mathcal{B}_{\mathbb{R}^3}$ y $\mathcal{B}_{\mathbb{R}^2}$ a las bases canónicas de $\mathbb{R}^3$ y $\mathbb{R}^2$, respectivamente, se tiene que

$$M_{\mathcal{B}_{\mathbb{R}^3}}^{\mathcal{B}_{\mathbb{R}^2}}(f) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Atendiendo al diagrama,

$$\begin{array}{ccccc} & f:\mathbb{R}^3 & \xrightarrow{\phantom{xxxx}} & \mathbb{R}^2 & \\ & & M_{\mathcal{B}_{\mathbb{R}^3}}^{\mathcal{B}_{\mathbb{R}^2}}(f) & & \\ & \mathcal{B}_{\mathbb{R}^3} & \longrightarrow & \mathcal{B}_{\mathbb{R}^2} & \\ M_{\mathcal{B}}^{\mathcal{B}_{\mathbb{R}^3}} \uparrow & & & & \downarrow M_{\mathcal{B}_{\mathbb{R}^2}}^{\mathcal{B}'} \\ & \mathcal{B} & \longrightarrow & \mathcal{B}' & \\ & & M_{\mathcal{B}}^{\mathcal{B}'}(f) & & \end{array}$$

obtendremos que $M_{\mathcal{B}}^{\mathcal{B}'}(f) = M_{\mathcal{B}_{\mathbb{R}^2}}^{\mathcal{B}'} M_{\mathcal{B}_{\mathbb{R}^3}}^{\mathcal{B}_{\mathbb{R}^2}}(f) M_{\mathcal{B}}^{\mathcal{B}_{\mathbb{R}^3}}$. Por tanto sólo tenemos que encontrar las matrices de cambio de base. Estas son:

$$M_{\mathcal{B}}^{\mathcal{B}_{\mathbb{R}^3}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad M_{\mathcal{B}_{\mathbb{R}^2}}^{\mathcal{B}'} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}$$

(recordemos que $M_{\mathcal{B}_{\mathbb{R}^2}}^{\mathcal{B}'} = (M_{\mathcal{B}'}^{\mathcal{B}_{\mathbb{R}^2}})^{-1}$), de manera que

$$M_{\mathcal{B}}^{\mathcal{B}'}(f) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \end{pmatrix}$$

(II) Sean $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2\}$ y $\mathcal{B}' = \{\mathbf{u}_1 + 2\mathbf{u}_2, 2\mathbf{u}_1 + 3\mathbf{u}_2\}$ bases de $\mathbb{R}^2$, y $f$ el endomorfismo cuya matriz respecto de la base $\mathcal{B}$ viene dada por

$$A = M_{\mathcal{B}}(f) = \begin{pmatrix} 6 & -2 \\ 6 & -1 \end{pmatrix}$$

La matriz de $f$ respecto de la base $\mathcal{B}'$ se deduce a partir del siguiente diagrama:

$$\begin{array}{ccc} f : \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ \mathcal{B} & \xrightarrow{A} & \mathcal{B} \\ C \uparrow & & \downarrow C^{-1} \\ \mathcal{B}' & \xrightarrow[A']{} & \mathcal{B}' \end{array}$$

donde $C = M_{\mathcal{B}'}^{\mathcal{B}}$ es la matriz de cambio de base de $\mathcal{B}'$ a $\mathcal{B}$. Recordemos que para obtener esta matriz debemos escribir por columnas, las coordenadas de los vectores de $\mathcal{B}'$ respecto de $\mathcal{B}$, es decir

$$C = M_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

Finalmente,

$$A' = M_{\mathcal{B}'}(f) = C^{-1}AC = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 6 & -2 \\ 6 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

**Ejemplo 5.9**

Dado el plano vectorial $V = x_1 - x_2 + 2x_3 = 0$, encontrar la matriz respecto de la base canónica de $\mathbb{R}^3$ de la simetría axial de eje $V$.

En este caso puede ser complicado calcular las imágenes de los vectores de la base canónica directamente, sin embargo, simples razonamientos geométricos nos pueden permitir calcular las imágenes de ciertos vectores de $\mathbb{R}^3$ que formen una base. Por ejemplo, la imagen por la simetría pedida de cualquier vector contenido en $V$ es el propio vector. Puesto que $V$ es un subespacio de dimensión dos, podemos extraer dos vectores l.i. sin más que resolver la ecuación dada. En este caso,

$$x_1 = x_2 - 2x_3 \Rightarrow \mathbf{u}_1 = (1, 1, 0),\ \ \mathbf{u}_2 = (-2, 0, 1) \in V$$

Obtener un tercer vector que forme base con los anteriores es bien sencillo, pero tal vez no sea fácil encontrar su imagen mediante la simetría. En tal caso, la mejor opción consistirá en escoger como tercer vector uno que, siendo independiente con los anteriores, su imagen sea sencilla de obtener. El candidato idóneo será un vector perpendicular a $V$, puesto que la simetría simplemente invierte el sentido de dicho vector (es decir, le cambia el signo).

Por geometría elemental se sabe que el vector formado por los coeficientes de la ecuación que define a $V$ es perpendicular a éste (en el tema 8 justificaremos este hecho), luego elegimos como tercer vector $\mathbf{u}_3 = (1, -1, 2)$. Tomando como base $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, es inmediato obtener la matriz de la simetría respecto de esta base, pues

$$f(\mathbf{u}_1) = \mathbf{u}_1, \quad f(\mathbf{u}_2) = \mathbf{u}_2, \quad f(\mathbf{u}_3) = -\mathbf{u}_3$$

La matriz queda

$$M_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Para hallar la matriz respecto de la base canónica simplemente usamos el cambio de base oportuno. En este caso

$$\begin{aligned} M_{\mathcal{B}_c}(f) &= \begin{pmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix} \end{aligned}$$

**Nota 5.4**

Como podemos deducir de (ii) del ejemplo anterior, si $V = W$, es decir, el espacio de partida y el de llegada son el mismo, y tomamos la misma base en ambos espacios, entonces $A' = C^{-1}AC$. Si ahora usamos el Teorema 2.16 es inmediato comprobar que $|A| = |A'|$. Es decir si $f \in \mathcal{L}(V)$ todas sus matrices asociadas poseen el mismo determinante.

## 5 5

## NÚCLEO Y RANGO DE UNA APLICACIÓN LINEAL

Dada una aplicación lineal existen un par de conjuntos asociados a partir de los cuales se pueden obtener de manera inmediata propiedades interesantes de la misma. Definimos estos conjuntos a continuación.

**Definición 5.2**

Sea $f : V \to W$ una aplicación. Se define el *núcleo* de $f$ como el conjunto

$$\ker(f) = \{\mathbf{v} \in V : f(\mathbf{v}) = \mathbf{0}\}$$

Se define la *imagen* de $f$ como

$$\operatorname{Im}(f) = \{\mathbf{w} \in W : \exists \mathbf{v} \in V \text{ tal que } f(\mathbf{v}) = \mathbf{w}\} = \{f(\mathbf{v}) : \mathbf{v} \in V\} = f(V)$$

Básicamente, el núcleo está constituido por aquellos vectores del espacio de partida que la aplicación transforma en el vector nulo. Por su parte, la imagen está formada por los vectores del espacio de llegada que son imagen de algún vector del espacio de partida.

**Proposición 5.5**

Sea $f : V \to W$ una aplicación lineal, entonces $\ker(f)$ y $\operatorname{Im}(f)$ son subespacios vectoriales de $V$ y $W$, respectivamente.

*Demostración:*
Usaremos la Proposición 4.9. Para ello, sean $\mathbf{u}$ y $\mathbf{v} \in \ker(f)$ y veamos que $\alpha\mathbf{u} + \beta\mathbf{v}$ también está en $\ker(f)$. En efecto,

$$f(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v}) = \mathbf{0}$$

Del mismo modo, si $\mathbf{u}, \mathbf{v} \in \text{Im}(f)$, veamos que $\alpha\mathbf{u} + \beta\mathbf{v}$ también está en $\text{Im}(f)$.

$$\left.\begin{array}{l} \mathbf{u} \in \text{Im}(f) \Rightarrow \exists \mathbf{x} \in V : f(\mathbf{x}) = \mathbf{u} \\ \mathbf{v} \in \text{Im}(f) \Rightarrow \exists \mathbf{y} \in V : f(\mathbf{y}) = \mathbf{v} \end{array}\right\}$$

Luego,

$$\alpha\mathbf{u} + \beta\mathbf{v} = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) = f(\alpha\mathbf{x} + \beta\mathbf{y})$$

Como $\alpha\mathbf{x} + \beta\mathbf{y} \in V$, se tiene el resultado deseado.

■

**Proposición 5.6**

Sea $f : V \to W$ una aplicación lineal. Entonces, $f$ es inyectiva si y sólo si $\ker(f) = \{\mathbf{0}\}$.

*Demostración:*

Supongamos que $f$ es inyectiva y que $\mathbf{x} \in \ker(f)$. Entonces $f(\mathbf{x}) = \mathbf{0} = f(\mathbf{0})$ (véase (I), Proposición 5.1). De la inyectividad deducimos que $\mathbf{x} = \mathbf{0}$, luego $\ker(f) = \{\mathbf{0}\}$.

Veamos la implicación contraria. Si $f(\mathbf{x}) = f(\mathbf{y})$ entonces $f(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Es decir, $\mathbf{x} - \mathbf{y} \in \ker(f)$. Como por hipótesis el núcleo se reduce al cero, $\mathbf{x} - \mathbf{y} = \mathbf{0}$, y por tanto $f$ es inyectiva.

■

**Ejemplo 5.10**

Sea $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ dada por $f(x_1, x_2) = (2x_1 + x_2, x_1 + 2x_2, 3x_1)$.
La matriz de $f$ en las bases canónicas de ambos espacios es

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 0 \end{pmatrix}$$

Para encontrar el núcleo buscamos los vectores de $\mathbb{R}^2$ (recordad que el núcleo está en el espacio de partida) tales que $f(x_1, x_2) = (0, 0, 0)$; es decir

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \left\{\begin{array}{rcl} 2x_1 + x_2 & = & 0 \\ x_1 + 2x_2 & = & 0 \\ 3x_1 & = & 0 \end{array}\right.$$

cuya única solución es el vector $(0, 0)$. Por tanto $f$ es inyectiva.

Es interesante observar que para encontrar los vectores del $\ker(f)$ hemos de resolver un sistema homogéneo cuya matriz coincide con la matriz de la aplicación lineal $f$. Sabemos que un sistema homogéneo tiene soluciones distintas de la solución trivial siempre que $\text{rango}(A) <$ nº de incógnitas. En el ejemplo anterior, puesto que $\text{rango}(A) = 2$, pudimos haber deducido automáticamente que $\ker(f) = \{\mathbf{0}\}$, sin necesidad de escribir ni resolver el sistema anterior, simplemente estudiando el rango de la matriz de la aplicación.

**Nota 5.5**

Si $A$ es la matriz de una aplicación lineal respecto de unas bases dadas, el sistema $A\mathbf{x} = \mathbf{0}$ es un sistema de ecuaciones implícitas de $\ker(A)$ y por tanto $\dim(\ker(A)) = n - \text{rango}(A)$, donde $n = \dim V$ (el espacio de partida).

**Proposición 5.7**

Sea $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ una base de $V$. Entonces el conjunto

$$\{f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)\}$$

es un sistema generador de $\text{Im}(f)$.

*Demostración:*
Sea $\mathbf{w} \in \text{Im}(f)$, entonces $\exists \mathbf{v} \in V$ tal que $f(\mathbf{v}) = \mathbf{w}$. Ahora bien, $\mathbf{v} \in V$ y $\mathcal{B}$ es una base de este espacio, por tanto

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i \Rightarrow \mathbf{w} = f(\mathbf{v}) = f\left(\sum_{i=1}^{n} \alpha_i \mathbf{e}_i\right) = \sum_{i=1}^{n} \alpha_i f(\mathbf{e}_i)$$

Es decir, hemos escrito un vector arbitrario de $\text{Im}(f)$ como combinación lineal de los vectores $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$. De aquí se obtiene el resultado.

■

**Nota 5.6**

Obsérvese que las columnas de la matriz de la aplicación $f$ son precisamente las imágenes de una base de $V$. Por tanto, los vectores cuyas coordenadas corresponden a las columnas de esta matriz forman un sistema generador de la $\text{Im}(f)$. Es decir $\dim(\text{Im}(f)) = \text{rango}(A)$.

En particular, todas las matrices de una misma aplicación lineal tienen el mismo rango, de ahí que hablemos del *rango de una aplicación*.

**Ejemplo 5.11**

Sea $f : \mathbb{R}^5 \to \mathbb{R}^3$ dada por

$$f(x_1, x_2, x_3, x_4, x_5) = (x_1 + x_2 + x_3 + x_4 + x_5, x_2 - x_4 + x_5, x_1 + x_3 + 2x_4)$$

Estudiar $\ker(f)$ y $\mathrm{Im}(f)$.

En primer lugar escribimos la matriz de $f$ en las bases canónicas:[3]

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 2 & 0 \end{pmatrix}$$

Un simple cálculo nos muestra que $\mathrm{rango}(A) = 2$ (un menor no nulo está formado por las primeras dos filas y columnas). Por tanto $\dim(\ker(f)) = 5 - 2 = 3$ y $\dim(\mathrm{Im}(f)) = 2$.

Para encontrar una base de $\mathrm{Im}(f)$ sólo tenemos que encontrar vectores l.i. de entre las columnas de $A$ (en este caso los correspondientes a las dos primeras columnas). Es decir,

$$\mathrm{Im}(f) = L((1, 0, 1), (1, 1, 0))$$

Para obtener una base de $\ker(f)$ resolvemos el sistema $A\mathbf{x} = \mathbf{0}$ (sólo consideramos las ecuaciones relevantes):

$$\left.\begin{array}{rcl} x_1 + x_2 + x_3 + x_4 + x_5 & = & 0 \\ x_2 - x_4 + x_5 & = & 0 \end{array}\right\} \Rightarrow \left.\begin{array}{rcl} x_1 + x_2 & = & -x_3 - x_4 - x_5 \\ x_2 & = & x_4 - x_5 \end{array}\right\}$$

Una base de soluciones es $\{(-1, 0, 1, 0, 0), (-2, 1, 0, 1, 0), (0, -1, 0, 0, 1)\}$.

[3] Siempre que no se especifique una base distinta escogeremos estas bases, pues son las más sencillas.

**Teorema 5.3**

Sea $f : V \to W$ lineal. Entonces

$$\dim(V) = \dim(\ker(f)) + \dim(\mathrm{Im}(f))$$

La demostración de este resultado es inmediata a partir de las notas 5.5 y 5.6. Como consecuencia se tiene el siguiente:

**Corolario 5.4**

Sea $f : V \to W$ lineal con matriz asociada $A$ respecto de ciertas bases. Entonces,

(I) $f$ es inyectiva si y sólo si $\mathrm{rango}(A) = \dim(V)$.

(II) $f$ es sobreyectiva si y sólo si $\mathrm{rango}(A) = \dim(W)$.

**Ejemplo 5.12**

Como consecuencia del resultado anterior, una aplicación $f : \mathbb{R}^n \to \mathbb{R}^m$ con $n > m$ no puede ser inyectiva, así como una aplicación de $\mathbb{R}^n$ en $\mathbb{R}^m$ con $n < m$ no puede ser sobreyectiva.

Resumimos los resultados anteriores en el siguiente teorema.

**Teorema 5.5**

Sea $f : V \to W$ lineal. Si $\dim(V) = \dim(W)$, son equivalentes:

(I) $f$ es biyectiva.

(II) $f$ es inyectiva.

(III) $f$ es sobreyectiva.

(IV) $\ker(f) = \{\mathbf{0}\}$.

(V) $\mathrm{rango}(f) = \dim(V)$.

5 6

## CÁLCULOS CON PYTHON

Prácticamente todos los conocimientos de Python que se necesitan para abordar los cálculos que se realizan en este tema han sido ya expuestos en temas anteriores (operaciones con matrices, resolución de sistemas y cálculo de rangos), sin embargo, aun no hemos hecho uso de una función de especial utilidad que nos va a permitir obtener el núcleo de una aplicación de forma sencilla: `nullspace`. Esta función, que forma parte del módulo SymPy, nos proporciona una base del $\ker(A)$, donde $A$ es una matriz cualquiera. Se usa del siguiente modo:

```
>>> from sympy import Matrix
>>> A=Matrix([[1,1,1,1,1],[0,1,0,-1,1],[1,0,1,2,0]])
>>> A.nullspace()
[[-1]
[ 0]
[ 1]
[ 0]
[ 0], [-2]
[ 1]
[ 0]
[ 1]
[ 0], [ 0]
[-1]
[ 0]
[ 0]
[ 1]]
```

Como vemos, la respuesta de este módulo, que funciona sobre una matriz de SymPy, nos proporciona una lista de vectores (en realidad son matrices columna) que corresponde a una base del espacio $\ker(A)$. De hecho, en este caso dicha base coincide con la obtenida por nosotros en el ejemplo 5.11, aunque es evidente que, en general, no tiene por qué ser así.

El interés de este módulo es que implícitamente nos está proporcionando el rango de la matriz, de manera que podemos definir ahora una función `rango` que trabaje sobre matrices de SymPy:

```
def rango(A):
        x=A.cols - len(A.nullspace())
        return x
```

El atributo `cols` hace referencia al número de columnas de $A$ (la dimensión del espacio de partida).

Además, `nullspace` también puede ser usado para calcular una base de un subespacio vectorial que venga definido por un sistema de ecuaciones homogéneo:

```
>>> from numpy import matrix
>>> from sympy import Matrix
>>> a=matrix('1 1 1 1; 0 0 1 1')
>>> A=Matrix(a)
>>> A.nullspace()
[[-1]
[ 1]
[ 0]
[ 0], [ 0]
[ 0]
[-1]
[ 1]]
```

El código anterior significa que el subespacio

$$L_1 \equiv \begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_3 + x_4 = 0 \end{cases}$$

tiene como base

$$\{(1, -1, 0, 0), (0, 0, -1, 1)\}$$

(véase el ejemplo 4.18).

Nótese que hemos usado `matrix` de NumPy para simplificar la entrada de la matriz, aunque luego hemos de usar `Matrix` de SymPy que es el objeto para el que está definido `nullspace`.

5 7

## APLICACIÓN A LA CRIPTOGRAFÍA

La *criptografía* es la técnica que altera una representación lingüística de un mensaje para mantener su privacidad. La información a transmitir es *encriptada* o *codificada* de forma que se haga ilegible hasta que sea *descodificada* o *desencriptada*. Para ello se emplea algún tipo de información secreta (la *clave* de encriptación) que debe ser conocida solo por el emisor y el receptor de la información.

Una de las técnicas criptográficas más sencillas consiste en enmascarar el mensaje original convirtiéndolo en un mensaje *cifrado* empleando para ello la *sustitución* de los caracteres que conforman el mensaje (por ejemplo, cambiar la *a* por la *k*, la *b* por la *d*, etc.). Pero este tipo de codificación es fácil de romper analizando la frecuencia de aparición de las letras del alfabeto. En esta sección vamos a mostrar una técnica criptográfica sencilla, pero más segura que la sustitución, que está basada en el uso de aplicaciones lineales.

Supongamos que queremos enviar el siguiente mensaje:

ALGEBRA LINEAL CON APLICACIONES Y PYTHON

En primer lugar vamos a sustituir los caracteres del mensaje por números, siguiendo un sencillo proceso de sustitución, para el que usamos la siguiente tabla:

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Ñ | O | P | Q | R | S | T | U | V | W | X | Y | Z | _ |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |

(5.4)

El mensaje codificado sería este:

1 12 7 5 2 19 1 28 12 9 14 5 1 12 28 3 16 14 28 1 17 12 9 3 1 3 9 16 14 5 20 28 26 28 17 26 21 8 16 14

Simplemente hemos sustituido cada letra por el número correspondiente, incluidos los espacios. Si enviáramos el mensaje de este modo, un sencillo análisis basándonos en la mayor frecuencia de aparición de ciertos números nos permitiría descifrar el mensaje. Sin embargo, vamos a codificar de nuevo el mensaje haciendo uso de una aplicación lineal biyectiva.

En primer lugar, agruparemos las cifras de las que consta el mensaje en grupos de 3:

$$(1,12,7)\ (5,2,19)\ (1,28,12)\ (9,14,5)\ (1,12,28)\ (3,16,14)\ (28,1,17)$$
$$(12,9,3)\ (1,3,9)\ (16,14,5)\ (20,28,26)\ (28,17,26)\ (21,8,16)\ (14,0,0)$$

Nótese que hemos añadido dos últimos números (0) para completar las ternas. Cada uno de estos grupos es un elemento de $\mathbb{R}^3$, de modo que vamos a transformarlos en un nuevo elemento de $\mathbb{R}^3$ mediante una aplicación lineal biyectiva, es decir, vamos a multiplicarlos por una matriz invertible. Para simplificar nuestros cálculos, elegiremos una matriz de enteros, cuya inversa también esté formada por enteros, por ejemplo:

$$A = \begin{pmatrix} 2 & 5 & 3 \\ 1 & 3 & 2 \\ 3 & 6 & 4 \end{pmatrix}$$

Ahora transformamos cada uno de los vectores mediante esta matriz; esto se puede hacer directamente así:

$$\begin{pmatrix} 2 & 5 & 3 \\ 1 & 3 & 2 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} 1 & 5 & 1 & 9 & 1 & 3 & 28 & 12 & 1 & 16 & 20 & 28 & 21 & 14 \\ 12 & 2 & 28 & 14 & 12 & 16 & 1 & 9 & 3 & 14 & 28 & 17 & 8 & 0 \\ 7 & 19 & 12 & 5 & 28 & 14 & 17 & 3 & 9 & 5 & 26 & 26 & 16 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 83 & 77 & 178 & 103 & 146 & 128 & 112 & 78 & 44 & 117 & 258 & 219 & 130 & 28 \\ 51 & 49 & 109 & 61 & 93 & 79 & 65 & 45 & 28 & 68 & 156 & 131 & 77 & 14 \\ 103 & 103 & 219 & 131 & 187 & 161 & 158 & 102 & 57 & 152 & 332 & 290 & 175 & 42 \end{pmatrix}$$

y entonces, el mensaje que enviamos es éste:

83 51 103 77 49 103 178 109 219 103 61 131 146 93 187 128 79 161 112 65 158
78 45 102 44 28 57 117 68 152 258 156 332 219 131 290 130 77 175 28 14 42

Obsérvese, por ejemplo, que la repetición de la cifra 28 en la primera sustitución podría ser una pista para determinar los espacios, sin embargo, ahora no existe tal repetición; de hecho, pocas son las cifras que se repiten.

Para descodificar el mensaje es fundamental que la aplicación usada sea biyectiva, pues debemos asegurarnos de que existe matriz inversa. Así, volviendo a agrupar el mensaje recibido en vectores tridimensionales:

$$(83,51,103)\ (77,49,103)\ (178,109,219)\ (103,61,131)\ (146,93,187)$$
$$(128,79,161)\ (112,65,158)\ (78,45,102)\ (44,28,57)\ (117,68,152)$$
$$(258,156,332)\ (219,131,290)\ (130,77,175)\ (28,14,42)$$

bastará multiplicar por $A^{-1}$, donde

$$A^{-1} = \begin{pmatrix} 0 & -2 & 1 \\ 2 & -1 & -1 \\ -3 & 3 & 1 \end{pmatrix}$$

para recuperar los dígitos de la sustitución:

$$\begin{pmatrix} 0 & -2 & 1 \\ 2 & -1 & -1 \\ -3 & 3 & 1 \end{pmatrix} \cdot$$
$$\begin{pmatrix} 83 & 77 & 178 & 103 & 146 & 128 & 112 & 78 & 44 & 117 & 258 & 219 & 130 & 28 \\ 51 & 49 & 109 & 61 & 93 & 79 & 65 & 45 & 28 & 68 & 156 & 131 & 77 & 14 \\ 103 & 103 & 219 & 131 & 187 & 161 & 158 & 102 & 57 & 152 & 332 & 290 & 175 & 42 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 5 & 1 & 9 & 1 & 3 & 28 & 12 & 1 & 16 & 20 & 28 & 21 & 14 \\ 12 & 2 & 28 & 14 & 12 & 16 & 1 & 9 & 3 & 14 & 28 & 17 & 8 & 0 \\ 7 & 19 & 12 & 5 & 28 & 14 & 17 & 3 & 9 & 5 & 26 & 26 & 16 & 0 \end{pmatrix}$$

Es sencillo implementar en Python funciones que codifiquen y descodifiquen mensajes, empleando el método anterior. Para facilitar la implementación hemos usado un tipo de dato no visto hasta el momento: los *diccionarios*.

Los *diccionarios* son una especie de arreglos, pero que en lugar de estar indexados por números, lo están por cualquier tipo de *clave*. Por ejemplo,

```
>>> dic={'Gauss':'1777','Grassmann':'1809','Hamilton
   ':'1805'}
```

Hemos definido un diccionario que vincula los nombres de tres matemáticos con su año de nacimiento. En este caso la clave es el nombre y el valor asociado a la clave el año de nacimiento. Podemos acceder a los elementos del diccionario como si fuera un arreglo:

```
>>> dic['Gauss']
'1777'
>>> dic['Hamilton']
'1805'
```

y podemos acceder a listas conteniendo las claves del siguiente modo:

```
>>> dic.keys()
['Grassmann', 'Hamilton', 'Gauss']
```

La principal ventaja de un diccionario es la facilidad con la que Python accede a sus elementos, lo que lo hace ventajoso para nuestros propósitos.

A continuación mostramos un código para encriptar y desencriptar mensajes:

```
#! /usr/bin/python
# -*- coding: UTF-8 -*-

from sys import exit
from numpy import array,dot

diccionario = {'a':'1','b':'2','c':'3','d':'4','e':'5','f
    ':'6','g':'7','h':'8','i':'9','j':'10','k':'11','l
    ':'12','m':'13','n':'14', '\xc3':'','\xb1':'15','o
    ':'16','p':'17','q':'18','r':'19','s':'20','t':'21','u
    ':'22','v':'23','w':'24','x':'25','y':'26','z':'27','
    ':'28'}

diccion = {'1':'a','2':'b','3':'c','4':'d','5':'e','6':'f
    ','7':'g','8':'h','9':'i','10':'j','11':'k','12':'l
    ','13':'m','14':'n','15':'\xc3\xb1','16':'o','17':'p
    ','18':'q','19':'r','20':'s','21':'t','22':'u','23':'v
    ','24':'w','25':'x','26':'y','27':'z','28':' ','0':''}

Mcod = array([[2,5,3],[1,3,2],[3,6,4]])
Mdecod = array([[0,-2,1],[2,-1,-1],[-3,3,1]])

def sustitucion(mensaje):

    mensus = []
    for v in mensaje:
        if v in diccionario.keys():
            if v!='\xc3':
                mensus.append(diccionario[v])
        else:
```

```
22            exit('El mensaje contiene caracteres no
                  permitidos')
23
24     if len(mensus)%3==0:
25         return mensus
26     elif len(mensus)%3==1:
27         mensus.append('0')
28         mensus.append('0')
29     else:
30         mensus.append('0')
31
32     return mensus
33
34
35 def desustit(lista):
36     texto=''
37     for v in lista:
38         texto += diccion[str(v)]
39     return texto
40
41
42 def crypt(lista,d=1):
43     n=len(lista)/3
44     a=array(lista,int)
45     a=a.reshape(3,n).T
46     if d==1:
47         res=dot(Mcod,a)
48     else:
49         res=dot(Mdecod,a)
50     res=res.T.reshape(3*n)
51     criptado=list(res)
52
53     return criptado
54
55
56 texto=raw_input()
57 mensaje=crypt(sustitucion(texto))
58 print "Este es el mensaje a enviar:"
59 print mensaje
60 print ' '
61 descod=desustit(crypt(mensaje,0))
62 print "Este es el mensaje descodificado:"
63 print descod
```

En la línea 7 definimos un diccionario que corresponde a la tabla de sustituciones (5.4). Hay un pequeño detalle concerniente a la letra *ñ*. Aunque hemos puesto la codificación correspondiente (línea 2), la lectura del carácter `ñ` en Python plantea algún problema técnico que hemos resulto definiendo en el dic-

cionario los caracteres `\xc3` y `\xb1`. Esto es debido a que la `ñ` es un carácter doble compuesto por `\xc3\xb1`.

Este diccionario se emplea en la sustitución inicial del mensaje de texto por una lista de números correspondientes a cada letra del mismo. Para realizar esta conversión definimos una función `sustitucion` (línea 14) cuya entrada es una cadena de caracteres, y cuya salida es una lista con los números correspondientes. La parte principal de la función consta de un bucle que recorre cada uno de los caracteres que conforma el texto entrante (línea 17).[4] Para cada carácter, se comprueba que éste está en el diccionario: si está, entonces se añade a la lista el elemento del diccionario correspondiente a ese carácter (línea 20), salvo que sea el primer carácter de la `ñ`, que se omite; y si no está, entonces el código se detiene pues no está preparado para admitir mensajes que contengan otro tipo de caracteres.[5] Obsérvese el uso del módulo `append` para añadir un elemento al final de una lista.

Por último, añadimos tantos caracteres `0` como necesitemos para que la longitud de la lista permita su agrupación en ternas (líneas 24–30). Nótese el uso de la división modular (resto de la división por un número) con el operador `%`.

Una vez convertido el texto a una lista de números hemos de codificarlo mediante la función `crypt` (línea 42), que esencialmente lo que hace es multiplicar por la matriz de codificación. Esta función admite como parámetro de entrada una lista y un valor entero `d`, que por defecto es igual a 1.

Básicamente, la función crea una matriz (en este caso, tipo *array* de NumPy) de tamaño $3 \times n$, donde $n$ es igual a la longitud de la lista dividida por 3. En la creación del arreglo en la línea 44 especificamos expresamente el tipo (entero) de los datos; lo hacemos así, pues la lista está formada por cadenas de texto y no números, como es necesario para poder multiplicar por la matriz.

Por otra parte, dado que la codificación y la descodificación suponen multiplicar por una matriz (o por su inversa) hemos usado la misma función para realizar las dos acciones. Para ello usamos el parámetro `d`, que si vale 1, esto es, el valor por defecto, entonces multiplicamos por la matriz `Mcod` (línea 47), mientras que si vale cualquier otra cosa, entonces multiplicamos por la matriz inversa `Mdecod` (línea 49). Ambas matrices han sido definidas al principio (líneas 11 y 12). Por último, la función redimensiona la matriz resultante para dar una lista.

Para finalizar, la función `desustit` (línea 35) realiza el proceso inverso a `sustitucion`: dada una lista de números, usa el diccionario `diccion` (definido en la línea 9) para restablecer el texto. En este caso, la función es más sencilla pues no hay que hacer comprobaciones. Solo señalar que puesto que los elementos de la lista ahora son números, hemos de usar la función `str` (línea 38) para convertirlos en cadenas de caracteres y así puedan ser identificadas como parte del diccionario.

El código anterior puede ser probado desde fuera del intérprete, pues hemos

[4] Python entiende las cadenas de texto como una especie de arreglo en el que sus caracteres se pueden recorrer uno a uno.

[5] Invitamos al lector a modificar el código para permitir otros caracteres (ejercicio 27).

añadido las líneas necesarias para proceder a la codificación y decodificación de un mensaje introducido por teclado. Para ello, en la línea 56 se solicita la entrada por teclado del mensaje mediante la función `raw_input()`, que se almacena en la variable `texto`. A continuación se sustituye y encripta el mensaje (línea 57), y luego se imprime para mostrar como queda. Finalmente, en la línea 61 el mismo mensaje es descodificado y deshecha la sustitución de números por letras, imprimiéndose éste. Obsérvese cómo en la llamada a la función `crypt` usamos un segundo argumento para indicar la codificación inversa.

Si guardamos el código anterior en un fichero con nombre `cripto.py`, la ejecución del código desde una consola sería ésta:

```
1 $ python cripto.py
2 teorema fundamental del algebra
3 Este es el mensaje a enviar:
4 [115, 68, 157, 102, 60, 139, 160, 97, 195, 126, 72, 166, 82,
      50, 101, 136, 79, 172, 176, 104, 220, 154, 97, 199, 83,
     51, 103, 77, 49, 103, 2, 1, 3]
5
6 Este es el mensaje descodificado:
7 teorema fundamental del algebra
```

donde la línea 2 es la que hemos introducido por teclado.

## 5.8 EJERCICIOS

### Ejercicios de repaso

**E.1** Decidir cuáles de las siguientes aplicaciones son lineales:

(a) $M_B : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathcal{M}_{2\times 1}(\mathbb{R})$ dada por $M_B(A) = AB$ con $B = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

(b) $N_B : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathcal{M}_{2\times 2}(\mathbb{R})$ dada por $N_B(A) = AB - BA$ con

$$B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

(c) $C : \mathcal{M}_n(\mathbb{R}) \to S$ donde $C(A) = \frac{1}{2}(A + A^T)$ y

$$S = \{A \in \mathcal{M}_n(\mathbb{R}) : A = A^T\}$$

**E.2** Escribir las matrices de las siguientes aplicaciones lineales respecto de las bases canónicas de los espacios correspondientes:

(a) $f(x_1, x_2, x_3) = (x_1 + x_2, 2x_1 - x_3)$

(b) $f(x_1, x_2, x_3) = (x_1 - x_2, x_2 + 2x_3, 2x_1 - 3x_3)$

(c) $f(x_1, x_2, x_3) = 3x_1 - x_2 + x_3$

**E.3** Dar una base para $\mathcal{M}_{3\times 2}(\mathbb{R})$ y hallar la matriz asociada a la aplicación lineal $T$ (aplicación traza) definida como $T : \mathcal{M}_{3\times 2}(\mathbb{R}) \to \mathbb{R}$ tal que $T(A) = \text{tr}(A)$, con respecto a esta base.

Nota: la traza de una matriz se definió en el ejercicio 22 del tema 4.

**E.4** Sea $T : \mathbb{P}^3_{\mathbb{R}}[x] \to \mathbb{P}^3_{\mathbb{R}}[x]$ lineal tal que $T(1) = x^2+1$, $T(x) = -x$, $T(x^2) = x^3$ y $T(x^3) = x^2 + x - 1$. Hallar la matriz asociada a $T$ con respecto a la base canónica.

**E.5** Se consideran las aplicaciones lineales $f, f' : \mathbb{R}^3 \to \mathbb{R}^2$ y $g, g' : \mathbb{R}^2 \to \mathbb{R}^3$ cuyas matrices respecto de las bases canónicas, son:

$$A = \begin{pmatrix} 2 & -1 & 3 \\ -1 & 0 & 3 \end{pmatrix} \quad A' = \begin{pmatrix} 1 & 2 & 4 \\ 2 & -1 & -3 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 0 \end{pmatrix} \quad B' = \begin{pmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 3 \end{pmatrix}$$

Se pide:

(a) Hallar las matrices respecto de las bases canónicas de:

$$(f + f') \circ g; \quad (f + f') \circ g'; \quad f \circ (g + g'); \quad g \circ (f + f');$$
$$g' \circ (f + f'); \quad (g + g') \circ f'; \quad (g + g') \circ (f + f'); \quad (f + f') \circ (g + g').$$

(b) Averiguar cuáles de las anteriores aplicaciones son inyectivos, cuáles sobreyectivos y cuáles isomorfismos.

**E.6** Demostrar que las aplicaciones $f : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $f(x, y) = (x + y, x + 2y)$ y $g : \mathbb{R}^2 \to \mathbb{R}^2$ dada por $g(x, y) = (2x - y, -x + y)$ son inversas una de la otra.

**E.7** Sea la aplicación lineal $T : \mathbb{R}^3 \to \mathbb{R}^3$ definida por

$$T(x, y, z) = (x + 2y + z, 3x + y + 2z, 3x + 2y + 2z)$$

Probar que es invertible y hallar la matriz asociada a $T^{-1}$.

**E.8** Sea $f$ el endomorfismo de $\mathbb{R}^3$ cuya matriz en la base $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ es

$$\begin{pmatrix} 1 & 2 & 0 \\ -1 & 2 & -1 \\ 3 & 0 & -4 \end{pmatrix}$$

Hallar la matriz del endomorfismo en la base $\mathcal{B}' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ siendo

$$\mathbf{v}_1 = \mathbf{u}_1, \quad \mathbf{v}_2 = -\mathbf{u}_2 + \mathbf{u}_1, \quad \mathbf{v}_3 = 2\mathbf{u}_1 - \mathbf{u}_2 + \tfrac{1}{2}\mathbf{u}_3$$

**E.9** Sea $V$ un espacio vectorial de dimensión 3 y $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ una base de $V$. Se considera la aplicación lineal $f : V \to V$ dada por:

$$f(\mathbf{u}_1) = -\mathbf{u}_1 - 2\mathbf{u}_2 + 6\mathbf{u}_3; \quad f(\mathbf{u}_2) = -\mathbf{u}_1 + 3\mathbf{u}_2; \quad f(\mathbf{u}_3) = -\mathbf{u}_1 - \mathbf{u}_2 + 4\mathbf{u}_3$$

Calcular las ecuaciones de $f$ respecto de las bases $\mathcal{B}'$ (en el espacio de partida) y $\mathcal{B}''$ (en el espacio de llegada), donde:

$$\mathcal{B}' = \{-\mathbf{u}_2 + \mathbf{u}_3, \mathbf{u}_3, -\mathbf{u}_1 - \mathbf{u}_2 + 4\mathbf{u}_3\}$$
$$\mathcal{B}'' = \{-\mathbf{u}_2 + \mathbf{u}_3, \mathbf{u}_3, -\mathbf{u}_1 - \mathbf{u}_2 + 3\mathbf{u}_3\}$$

**E.10** Dadas las siguientes aplicaciones lineales encontrar una base del núcleo y unas ecuaciones implícitas de la imagen comprobando en cada caso la ecuación de dimensiones:

$$\dim(\ker(A)) + \dim(\operatorname{Im}(A)) = \dim(V)$$

(donde $V$ es el espacio de partida) e indicar si son inyectivas o sobreyectivas:

(a) $A : \mathbb{R}^3 \to \mathbb{R}^3$ definida por $A(x, y, z) = (x + z, y, x + y + z)$.

(b) $B : \mathbb{C}^2 \to \mathbb{C}^2$ definida por $B(z_1, z_2) = (iz_1 + z_2, z_1 + iz_2)$.

(c) $D : \mathbb{C}^3 \to \mathbb{C}^3$ con matriz asociada

$$D = \begin{pmatrix} 1 & i & 0 \\ 0 & 1 & 2 \\ i & -1 & 0 \end{pmatrix}$$

## Problemas

**E.11** Encontrar las matrices de las siguientes aplicaciones lineales:

(a) $M_B$ y $N_B$ del ejercicio 1.

(b) $A : \mathbb{P}^2_{\mathbb{R}}[x] \to \mathbb{R}^3$ tal que $A(p(x)) = (p(0), p(1), p(2))$.

(c) $A : \mathbb{P}^3_{\mathbb{R}}[x] \to \mathbb{P}^3_{\mathbb{R}}[x]$ tal que $A(p(x)) = p(x + 1)$.

**E.12** Sea $A : \mathbb{R}^3 \to \mathbb{R}^3$ dada por $A(x, y, z) = (x + y, z, x + z)$. Encontrar la matriz asociada a $A$ con respecto a la base canónica de $\mathbb{R}^3$. Hallar la imagen mediante $A$ de los siguientes subespacios vectoriales:

(a) $V_1 = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$.

(b) $V_2 = \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) = t(1, -1, 1)\ t \in \mathbb{R}\}$.

**E.13** Sea $f$ un endomorfismo de $\mathbb{R}^2$ tal que $f(1,0) = (2,4)$ y $f(0,1) = (5,-1)$. Hallar la imagen por $f$ del vector $(2,3)$ y la imagen inversa por $f$ del vector $(0,3)$.

**E.14** Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ una aplicación lineal tal que $f(1,1) = (0,-1)$ y $f(-1,2) = (3,1)$. ¿Es $f$ invertible? En caso afirmativo calcular $f^{-1}(1,1)$.

**E.15** Respecto de la base canónica de $\mathbb{R}^3$ hallar las matrices de las siguientes aplicaciones lineales:

(a) Simetría con respecto a la recta vectorial $\{x = 0,\ y = 0\}$.

(b) Simetría con respecto a la recta vectorial $(x, y, z) = t(1, 1, 1)$.

(c) Simetría con respecto al plano vectorial $x = y$.

(d) Giro de $\alpha$ radianes con respecto al eje $OZ$.

(e) Giro de 90 grados con respecto a la recta $\{x + y = 0,\ z = 0\}$

(f) Proyección sobre el plano vectorial $x - y + z = 0$.

**E.16** Sea $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ la aplicación lineal definida por

$$T(x_1, x_2) = (x_1 + 2x_2, -x_1, 0)$$

(a) Encontrar la matriz de la aplicación $T$ en las bases:

$$\mathcal{B}_1 = \{\mathbf{u}_1 = (1,3),\ \mathbf{u}_2 = (-2,4)\} \text{ y}$$
$$\mathcal{B}_2 = \{\mathbf{v}_1 = (1,1,1),\ \mathbf{v}_2 = (2,2,0),\ \mathbf{v}_3 = (3,0,0)\}$$

(b) Usar la matriz obtenida en el apartado anterior para calcular $T(8,3)$.

**E.17** Sea $f : \mathbb{R}^3 \to \mathbb{R}^3$ una aplicación lineal tal que $\ker(f) = \{\mathbf{x} \in \mathbb{R}^3 : 2x_1 + x_2 + x_3 = 0\}$, Sea $\mathbf{u} = (1,-1,2)$ y $\mathbf{v} = (-2,2,\alpha)$. ¿Para qué valor de $\alpha$ se cumple que $f(\mathbf{u}) = f(\mathbf{v})$.

**E.18** Sea $f : \mathbb{R}^4 \to \mathbb{R}^3$ una aplicación lineal tal que:

- $\ker(f) = \{\mathbf{x} \in \mathbb{R}^4 : 5x_1 - x_2 - 5x_3 = 0,\ 5x_1 + x_2 - 5x_4 = 0\}$
- $\mathrm{Im}(f) = \{\mathbf{y} \in \mathbb{R}^3 : y_1 + y_2 - y_3 = 0\}$
- $f(1,2,0,1) = (1,1,2)$
- Existe $\lambda \in \mathbb{R}$ tal que $f(0,1,2,1) = (\lambda, \lambda + 1, \lambda)$

Calcular $f(3,-1,-3,0)$.

**E.19** Sea $V$ un espacio vectorial de dimensión 4 y $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ una base de $V$. Sea $f$ el endomorfismo de $V$ dado por

$$f(\mathbf{u}_1) = \mathbf{u}_1 - 3\mathbf{u}_2 + 2\mathbf{u}_3 + 2\mathbf{u}_4$$
$$f(\mathbf{u}_2) = -2\mathbf{u}_2 + \mathbf{u}_3 + 3\mathbf{u}_4$$
$$f(\mathbf{u}_3) = 4\mathbf{u}_1 - 2\mathbf{u}_2 + 3\mathbf{u}_3 - 7\mathbf{u}_4$$
$$f(\mathbf{u}_4) = 4\mathbf{u}_1 - 6\mathbf{u}_2 + 5\mathbf{u}_3 - \mathbf{u}_4$$

Se pide:

(a) Escribir las ecuaciones de $f$ respecto de la base $\mathcal{B}$, y hallar una base y la dimensión de $\ker(f)$ e $\operatorname{Im}(f)$.

(b) Sea $L$ la variedad lineal generada por los vectores

$$\mathbf{u}_1 - \mathbf{u}_2 + 2\mathbf{u}_4, \quad -3\mathbf{u}_1 + 4\mathbf{u}_3 + 2\mathbf{u}_4, \quad -3\mathbf{u}_1 + 2\mathbf{u}_2 + 3\mathbf{u}_4$$

Calcular una base y unas ecuaciones implícitas de $f(L)$ respecto de la base $\mathcal{B}$.

**E.20** Dar la dimensión y una base del núcleo y la imagen de las siguientes aplicaciones lineales, indicando si son inyectivas, sobreyectivas y/ó biyectivas:

(a) $M_B$ y $N_B$ del ejercicio 1.

(b) $A : \mathbb{P}^3_{\mathbb{R}}[x] \to \mathbb{P}^3_{\mathbb{R}}[x]$ tal que $A(1) = x^2+1$, $A(x) = x+2$, $A(x^2) = x^3 - x$ y $A(x^3) = 1$.

(c) La aplicación derivada de $\mathbb{P}^3_{\mathbb{R}}[x]$ en $\mathbb{P}^2_{\mathbb{R}}[x]$.

(d) La aplicación traza en $\mathcal{M}_2(\mathbb{R})$.

**E.21** Sea $f$ el endomorfismo de $\mathcal{M}_2(\mathbb{R})$ definido por $f(X) = A \cdot X - X \cdot A$, donde $A$ es la matriz

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Se pide:

(a) Calcular las ecuaciones de $f$ respecto de la base canónica de $\mathcal{M}_2(\mathbb{R})$

(b) Probar que $V = \operatorname{Im}(f) \oplus \ker(f)$.

### Ejercicios teóricos

**E.22** Sea $f : V \to V$ un endomorfismo tal que $f^2 + f + \mathrm{Id} = \mathbf{0}$. ¿Es $f$ invertible?

***E.23** Dada $A \in \mathcal{L}(V)$ demostrar:

(a) $A^2 = 0$ si y sólo si $\mathrm{Im}(A) \subset \ker(A)$.

(b) $\ker(A^{n-1}) \subset \ker(A^n)$, para $n \geq 1$.

(c) Si $V = \mathrm{Im}(A) \oplus \ker(A)$ entonces $\mathrm{Im}(A) = \mathrm{Im}(A^2)$.

**E.24** Sea $f : V \to V'$ una aplicación lineal entre los e.v. $V$ y $V'$ y sea $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ una base de $V$. Probar que:

(a) $f$ es inyectivo si y sólo si $\{f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)\}$ es l.i.

(b) $f$ es sobreyectiva si y sólo si $\{f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)\}$ es un sistema de generadores de $V'$.

(c) $f$ es biyectivo si y sólo si $\{f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)\}$ es una base de $V'$.

**E.25** Sean $f$ y $g$ dos endomorfismos de un espacio vectorial tales que $g \circ f = \mathbf{0}$. Decidir cuáles de las siguientes afirmaciones son ciertas:

(a) $\mathrm{Im}(f) \subset \ker(g)$.

(b) $\mathrm{Im}(g) \subset \ker(f)$.

(c) $\ker(f) \subset \mathrm{Im}(g)$.

(d) $\ker(g) \subset \mathrm{Im}(f)$.

**E.26** Sea $f$ y $g$ dos endomorfismos de un espacio vectorial tales que $\dim(\ker(f)) = 1$ y $\dim(\ker(g \circ f)) = 1$. Calcular $\dim(\ker(g))$.

### Ejercicios adicionales

**E.27** Modificar el código de la página 218 para que los mensajes puedan llevar caracteres mayúsculas, números y caracteres acentuados.

# 6 Diagonalización

En el tema anterior hemos estudiado las matrices asociadas a una aplicación y hemos visto cómo éstas cambian en función de la base elegida para representarla. El objetivo principal de este tema es tratar de obtener una matriz que represente a un endomorfismo dado, respecto de cierta base, que sea lo más sencilla posible. Como veremos, esta representación nos proporcionará información interesante sobre el funcionamiento de una aplicación lineal.

Para entender mejor este objetivo retomemos (II) del ejemplo 5.8 en el que teníamos un endomorfismo $A$ cuya matriz respecto de cierta base $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2\}$ era

$$A = \begin{pmatrix} 6 & -2 \\ 6 & -1 \end{pmatrix}$$

mientras que en la base $\mathcal{B}' = \{\mathbf{v}_1, \mathbf{v}_2\}$, con

$$\mathbf{v}_1 = \mathbf{u}_1 + 2\mathbf{u}_2, \quad \mathbf{v}_2 = 2\mathbf{u}_1 + 3\mathbf{u}_2$$

correspondía a

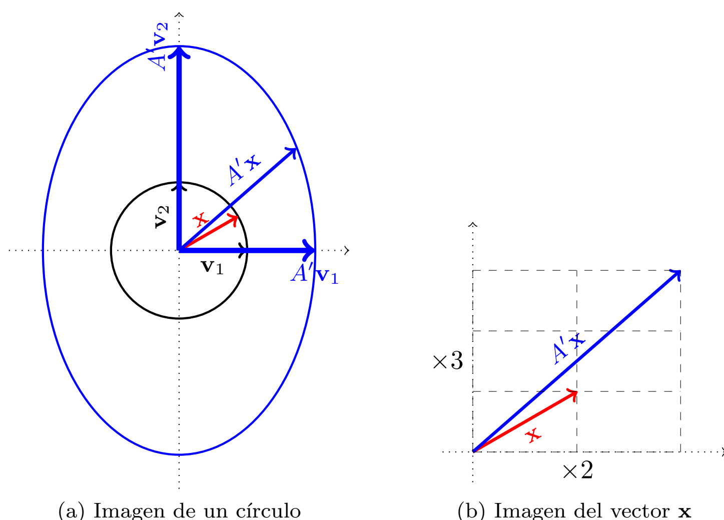
$$A' = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Es decir, en esta situación tenemos una base respecto de la cual, la aplicación lineal dada tiene asociada una matriz diagonal. Tales matrices tienen interesantes ventajas; por una parte porque son fáciles de operar (son fáciles de multiplicar, de invertir, etc.) pero además nos proporcionan una mejor comprensión de lo que hace la aplicación.

Más concretamente, si prestamos atención a la aplicación anterior respecto de la base $\mathcal{B}'$, atendiendo a la definición de matriz de una aplicación respecto de una base, observamos que

$$A'\mathbf{v}_1 = 2\mathbf{v}_1, \quad A'\mathbf{v}_2 = 3\mathbf{v}_2$$

es decir, la imagen de cada uno de los vectores de la base elegida corresponde a un múltiplo de ellos mismos. Desde el punto de vista geométrico es fácil hacerse una idea de lo que hace esta aplicación. En la figura 6.1a vemos una muestra de ello. Suponiendo que los vectores $\mathbf{v}_1$ y $\mathbf{v}_2$ son los vectores de los ejes, esta aplicación "estira" los vectores en la dirección $\mathbf{v}_1$ hasta el doble de su tamaño, y




(a) Imagen de un círculo

(b) Imagen del vector $\mathbf{x}$

Figura 6.1: Aplicación lineal $A'$

los de la dirección $\mathbf{v}_2$ al triple de su longitud. De modo que un círculo centrado en el origen se transformará en un elipse de semiejes 2 y 3.

En la figura 6.1b vemos con más detalles cómo se transforma un vector $\mathbf{x}$ cualquiera: descomponemos el vector en sus componentes horizontal y vertical, y la imagen tendrá su componente horizontal multiplicada por 2 y la vertical por 3.

Obsérvese que el hecho de que cada uno de los vectores de la base se transforme en un múltiplo suyo (es decir, esencialmente en el mismo vector) es lo que nos ha permitido describir geométricamente la aplicación. Estos vectores *especiales* son los que tratamos en la siguiente sección.

## 6 1 VALORES Y VECTORES PROPIAS

Los autovalores y autovectores juegan un papel muy importante a la hora de entender el comportamiento de los endomorfismos, lo cual se pondrá de manifiesto principalmente en el siguiente tema, en el que su uso será fundamental.[1]

[1] Si bien estos conceptos están asociados a aplicaciones lineales, fueron descubiertos primero en el estudio de formas cuadráticas y de ecuaciones diferenciales a través de los trabajos de Euler, Lagrange y Cauchy entre finales del s. XVIII y principios del XIX. Fue el matemático alemán David Hilbert, a comienzos del s. XX, quien introdujo por primera vez los términos *eigenvalue* y *eigenvector* (autovalor y autovector, respectivamente).

**Definición 6.1**

Un vector $\mathbf{x} \neq \mathbf{0}$ se dice *vector propio* o *autovector* de un endomorfismo $f : V \to V$ si existe $\lambda \in \mathbb{K}$ tal que $f(\mathbf{x}) = \lambda\mathbf{x}$. A $\lambda$ se le llama *valor propio* o *autovalor* asociado al autovector $\mathbf{x}$.

Es decir, los autovectores son vectores que se transforman "esencialmente" en sí mismos (salvo constantes, que son los autovalores).

**Nota 6.1**

(I) Es importante hacer hincapié en que la definición impide que el vector nulo sea autovector, sin embargo sí es posible que una aplicación tenga como autovalor 0.

(II) Si consideramos $\mathbf{x}$ un autovector de una aplicación $f$ asociado a cierto autovalor $\lambda$, y $L = L(\mathbf{x})$ es el subespacio que genera dicho vector, entonces cualquier otro vector $\mathbf{y} \in L$ también es autovector asociado al mismo autovalor. Es decir, hay infinitos autovectores asociados a un autovalor.

Esto es fácil de comprobar, puesto que si $\mathbf{y} \in L$ entonces $\mathbf{y}$ es de la forma $\mathbf{y} = \alpha\mathbf{x}$, para cierto $\alpha \in \mathbb{K}$; luego $f(\mathbf{y}) = \alpha f(\mathbf{x}) = \alpha\lambda\mathbf{x} = \lambda\mathbf{y}$, pues $\mathbf{x}$ es autovector.

Podemos observar que todos los vectores de $L$ permanecen en $L$. Este es el concepto de invarianza que presentamos a continuación.

Otro concepto fuertemente vinculado a los autovectores y autovalores es el de *subespacio invariante*.

**Definición 6.2**

Sea $f : V \to V$ una aplicación lineal. Un conjunto $L$ se dice *invariante* respecto de $f$ si $f(L) \subset L$.

**Ejemplo 6.1**

(I) Como hemos comentado en la nota 6.1, si $f$ es la aplicación lineal cuya matriz en la base $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ es

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

entonces los subespacios $L_1 = L(\mathbf{v}_1)$ y $L_2 = L(\mathbf{v}_2)$ son invariantes respecto de $f$.

(II) Dada una matriz de rotación respecto del eje $OZ$ de ángulo $\theta$, cuya matriz respecto de la base canónica se escribe

$$R = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta & 0 \\ \operatorname{sen}\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

es fácil deducir geométricamente sus invariantes: el plano $L_1 = \{x_3 = 0\}$ y el eje $OZ$ ($L_2 = \{x_1 = x_2 = 0\}$).

Comprobémoslo: si $\mathbf{x} \in L_1 \Rightarrow \mathbf{x} = \alpha(1, 0, 0) + \beta(0, 1, 0)$. Entonces,

$$\begin{aligned} R\mathbf{x} &= \alpha(\cos\theta, \operatorname{sen}\theta, 0) + \beta(-\operatorname{sen}\theta, \cos\theta, 0) \\ &= (\alpha\cos\theta - \beta\operatorname{sen}\theta, \alpha\operatorname{sen}\theta + \beta\cos\theta, 0) \in L_1 \end{aligned}$$

Mientras que si $\mathbf{x} \in L_2$ se observa que $R\mathbf{x} = \mathbf{x}$. Nótese que esta igualdad significa que los vectores de $L_2$ son autovectores asociados al autovalor 1.

(III) Si $f : V \to V$ es lineal, los espacios $\{\mathbf{0}\}$ y $V$ son trivialmente invariantes para dicha aplicación.

Los espacios invariantes tienen un especial significado desde un punto de vista geométrico que ayuda a describir el comportamiento de una aplicación. Como hemos podido comprobar en el ejemplo anterior, estos espacios están íntimamente relacionados con los vectores propios. De momento probaremos un sencillo resultado sobre operaciones con subespacios invariantes.

**Proposición 6.1**

Si $L_1$ y $L_2$ son subespacios invariantes de una aplicación lineal $f$ entonces $L_1 + L_2$ y $L_1 \cap L_2$ también son subespacios invariantes.

*Demostración:*
Veámoslo para la suma. Si $\mathbf{x} \in L_1 + L_2$ entonces por definición de suma de subespacios $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ con $\mathbf{x}_1 \in L_1$ y $\mathbf{x}_2 \in L_2$. Luego $f(\mathbf{x}) = f(\mathbf{x}_1) + f(\mathbf{x}_2)$ pertenece a $L_1 + L_2$, pues $f(\mathbf{x}_1) \in L_1$ y $f(\mathbf{x}_2) \in L_2$, ya que ambos subespacios son invariantes.

Análogamente se razona para la intersección de subespacios.

■

Como comentamos al inicio del tema, el objetivo que perseguimos es el de poder representar una aplicación lineal de la forma más sencilla posible. En el ejemplo que mostramos al inicio hemos podido comprobar que es fácil interpretar una aplicación lineal cuando la matriz de ésta es diagonal. La siguiente definición va en esa línea.

**Definición 6.3**

Una aplicación lineal $f : V \to V$ se dice *diagonalizable* si existe una base de $V$ respecto de la cual la matriz asociada es diagonal.

Igualmente, se dice que una matriz es diagonalizable si la aplicación lineal que la tiene como matriz asociada es diagonalizable. Esto es, si existe $C \in \mathcal{M}_n(\mathbb{K})$ invertible tal que $D = C^{-1}AC$, con $D$ una matriz diagonal.[2]

El siguiente resultado nos proporciona una caracterización de las aplicaciones diagonalizables:

**Teorema 6.1**

Una aplicación lineal $f : V \to V$ es diagonalizable si y sólo si existe una base de $V$ formada por autovectores de $f$.

*Demostración:*
Probemos en primer lugar la suficiencia. Si $\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ es una base

[2] Esto último se tiene al realizar un cambio de base, desde la base respecto de la que tenemos la matriz de la aplicación a la base respecto de la cual la aplicación tiene una matriz asociada que es diagonal. Como vimos en el tema anterior, la relación entre ambas matrices es de la forma $D = C^{-1}AC$.

formada por autovectores de $f$, con autovalores asociados $\lambda_1, \dots, \lambda_n$, entonces por definición, $f(\mathbf{e}_j) = \lambda_j \mathbf{e}_j$. Es decir, las coordenadas respecto de la base $\mathcal{B}$ de la imagen del vector $\mathbf{e}_j$ son $(0, \dots, 0, \lambda_j, 0, \dots, 0)$. Por tanto, la matriz respecto de esta base se escribe

$$M_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{6.1}$$

que es diagonal.

Veamos ahora la necesidad. Si la matriz respecto de una base $\mathcal{B}' = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ es diagonal tendrá la forma dada en (6.1). Es decir,

$$f(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1, \quad \dots, \quad f(\mathbf{v}_n) = \lambda_n \mathbf{v}_n,$$

es decir, existe una base formada por autovectores.

■

La pregunta natural que surge es: ¿siempre es posible obtener esto? Es decir, dada una aplicación, ¿podemos encontrar siempre una base respecto de la cual la matriz de dicha aplicación sea diagonal? Veamos que la respuesta es negativa a través del siguiente ejemplo:

**Ejemplo 6.2**

Consideremos la aplicación lineal cuya matriz respecto de cierta base sea

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Si existe una base respecto de la cual dicha aplicación tiene asociada una matriz diagonal, entonces, usando las matrices de cambio de base se debería tener:

$$C^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} C = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

para cierta matriz invertible $C$, que podemos escribir como

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Haciendo operaciones,

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \frac{1}{|C|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{|C|} \begin{pmatrix} ad\alpha - bc\beta & -ab\alpha + ab\beta \\ cd\alpha - cd\beta & -bc\alpha + ad\beta \end{pmatrix}$$

Por tanto hemos de resolver el sistema

$$\left.\begin{array}{rcll} \frac{1}{|C|}(ad\alpha - bc\beta) & = & 1 & (1) \\ \frac{1}{|C|}ab(-\alpha+\beta) & = & 1 & (2) \\ \frac{1}{|C|}cd(\alpha-\beta) & = & 0 & (3) \\ \frac{1}{|C|}(-bc\alpha + ad\beta) & = & 1 & (4) \end{array}\right\}$$

Como $|C| \neq 0$ (pues $C$ es invertible), de (3) se deduce que:

(I) O bien $c = 0$, en cuyo caso de (1) y (2) se tiene $\frac{1}{|C|}ad\alpha = \frac{1}{|C|}ad\beta$. De aquí, si $a = 0$, (2) no es posible. Si $d = 0$, (1) no se tendría, y la otra opción es que $\alpha = \beta$, pero esto contradice nuevamente (2). Por tanto $c \neq 0$.

(II) Si $d = 0$ se tendría que $-\frac{1}{|C|}bc\beta = -\frac{1}{|C|}bc\alpha$, y un argumento similar al anterior nos lleva a contradicción. Luego $d \neq 0$.

(III) La única posibilidad que resta es que $\alpha = \beta$, que contradice (2).

En conclusión no puede existir $C$ invertible tal que $C^{-1}BC$ sea diagonal, o dicho de otro modo, no toda aplicación es diagonalizable.

En lo que sigue trataremos de dar resultados que nos garanticen que una aplicación es diagonalizable, y finalmente veremos cómo podemos obtener una base respecto de la cual, la matriz de una aplicación sea lo más sencilla posible (aunque no sea diagonal).

**Teorema 6.2**

Si $\mathbf{e}_1, \ldots \mathbf{e}_n$ son autovectores correspondientes a autovalores $\lambda_1, \ldots, \lambda_n$ distintos entre sí dos a dos, entonces $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ es un conjunto l.i.

*Demostración:*

Procederemos por inducción en el número de vectores. Para $n = 1$ el resultado es trivial (recuérdese que un vector no nulo siempre es l.i. y que los autovectores, por definición, no pueden ser nulos).

Supuesto el resultado cierto para $n-1$, veamos que también es cierto para $n$. Para probar que $n$ vectores son l.i. consideramos una combinación lineal igualada al vector nulo y trataremos de ver que todos los escalares deben ser

iguales a cero:

$$\begin{aligned} \alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n = \mathbf{0} \quad &\Leftrightarrow \quad f(\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n) = f(\mathbf{0}) = \mathbf{0} \\ &\Leftrightarrow \quad \alpha_1 f(\mathbf{e}_1) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0} \\ &\Leftrightarrow \quad \alpha_1\lambda_1\mathbf{e}_1 + \cdots + \alpha_n\lambda_n\mathbf{e}_n = \mathbf{0} \qquad (6.2) \end{aligned}$$

Multiplicando $\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n = \mathbf{0}$ por $\lambda_n$ y restándola de (6.2) se obtiene

$$\alpha_1(\lambda_n - \lambda_1)\mathbf{e}_1 + \cdots + \alpha_{n-1}(\lambda_n - \lambda_{n-1})\mathbf{e}_{n-1} = \mathbf{0}$$

Esta expresión es una combinación de los $n-1$ primeros vectores igualada a cero. Usando la hipótesis de inducción, estos vectores son l.i., y por tanto los escalares deben ser nulos; pero como además $\lambda_n - \lambda_j \neq 0$, $\forall j$ (pues por hipótesis los autovalores son todos distintos), se tendrá que $\alpha_1 = \cdots = \alpha_{n-1} = 0$. Sustituyendo en la combinación lineal inicial se deduce que $\alpha_n$ también es cero, y de aquí se obtiene la independencia buscada.

■

**Nota 6.2**

Como consecuencia de los Teoremas 6.1 y 6.2, si un endomorfismo en un espacio de dimensión $n$ tiene $n$ autovalores distintos entre sí, entonces será diagonalizable. Como veremos más adelante, el recíproco no es cierto, es decir, existen endomorfismos en espacios de dimensión $n$ diagonalizables, que no tienen $n$ autovalores distintos.

6 2

## POLINOMIO CARACTERÍSTICO

Veamos ahora cómo podemos calcular autovectores y autovalores de una aplicación. Sea $f : V \to V$ una aplicación lineal con matriz asociada $A$ en una cierta base. Sabemos que

$$f(\mathbf{x}) = A\mathbf{x}$$

Si $\mathbf{x}$ es un autovector asociado a un autovalor $\lambda$, entonces

$$f(\mathbf{x}) = \lambda\mathbf{x} \Rightarrow A\mathbf{x} = \lambda\mathbf{x} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}$$

Denotando por $A = (a_{ij})$ y $\mathbf{x} = (x_1, \ldots, x_n)$, podemos escribir estas ecuaciones explícitamente como

$$\left.\begin{array}{ccccccccc} (a_{11}-\lambda)x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & (a_{22}-\lambda)x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ & & & & & & & \vdots & \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & (a_{nn}-\lambda)x_n & = & 0 \end{array}\right\}$$

obteniendo un sistema homogéneo que debe satisfacer cada uno de los autovectores asociados al autovalor $\lambda$. Así pues, para que existan autovectores, este sistema debe tener solución distinta de la trivial, o equivalentemente, la matriz del sistema debe tener rango menor que $n$. Esto es,

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0 \Leftrightarrow |A-\lambda I| = 0$$

**Definición 6.4**

Dada $A \in \mathcal{M}_n(\mathbb{K})$, la expresión $|A-\lambda I|$ es un polinomio[3] en la variable $\lambda$ que se denomina *polinomio característico*[4] de $A$.

**Nota 6.3**

Como hemos visto antes, los autovalores de una aplicación lineal son las raíces del polinomio característico. Por tanto, si queremos asegurar la existencia de autovalores necesitamos que el cuerpo $\mathbb{K}$ sea algebraicamente cerrado, es decir, debemos trabajar en $\mathbb{C}$, de forma que la ecuación $|A-\lambda I| = 0$ tenga $n$ soluciones (contando multiplicidades).

Como hemos visto, el polinomio característico se construye usando la matriz de una aplicación en una base dada. Pero ¿qué ocurre si cambiamos la base? El siguiente resultado nos muestra que el polinomio característico no cambia si usamos otra base para representar a la matriz de una aplicación.

**Teorema 6.3**

El polinomio característico no depende de la base que representa a la matriz de una aplicación.

*Demostración:*
Sean $A$ y $A'$ dos matrices asociadas a la misma aplicación lineal en las bases $\mathcal{B}$

[3]Se puede probar por inducción que dicha expresión es de hecho un polinomio de grado exactamente $n$.

[4]El concepto aparece explícitamente en un trabajo de Lagrange de 1774 relacionado con ecuaciones diferenciales.

y $\mathcal{B}'$, respectivamente. Los polinomios característicos serán

$$\mathcal{P}_{\mathcal{B}}(\lambda) = |A - \lambda I|, \quad \mathcal{P}_{\mathcal{B}'}(\lambda) = |A' - \lambda I|$$

Sabemos que la relación entre las matrices $A$ y $A'$ viene dada por $A' = C^{-1}AC$, con $C$ una matriz invertible (la matriz del cambio de base). Entonces,

$$\begin{aligned} |A' - \lambda I| &= |C^{-1}AC - \lambda I| = |C^{-1}AC - \lambda C^{-1}IC| \\ &= |C^{-1}AC - C^{-1}(\lambda I)C| = |C^{-1}(AC - \lambda IC)| \\ &= |C^{-1}(A - \lambda I)C| = |C^{-1}||A - \lambda I||C| = |A - \lambda I| \end{aligned}$$

es decir, ambos polinomios son iguales.

■

**Ejemplo 6.3**

(I) Consideremos la aplicación de matriz siguiente:

$$A = \begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}$$

El polinomio característico vendrá dado por:

$$\mathcal{P}_A(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 5 & 4-\lambda \end{vmatrix} = (1-\lambda)(4-\lambda) - 10 = \lambda^2 - 5\lambda - 6$$

Sus raíces, esto es, los autovalores de $A$, son $\lambda = -1$ y $\lambda = 6$.

Según comentamos en la nota 6.2 al tener todos los autovalores distintos, la aplicación será diagonalizable, pues los autovectores asociados deben ser l.i., y como estamos en un espacio de dimensión dos, estos autovectores forman una base.

Calculemos ahora los autovectores. Debemos resolver la ecuación $A\mathbf{x} = \lambda\mathbf{x}$ para cada autovalor obtenido. Esto es, $(A - \lambda I)\mathbf{x} = \mathbf{0}$;

para $\lambda = -1$,

$$(A + I)\mathbf{x} = \mathbf{0} \Rightarrow \begin{pmatrix} 2 & 2 \\ 5 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Nótese que este sistema ha de tener soluciones no triviales, pues en caso contrario no habría autovectores, lo cual no puede ocurrir pues

los autovalores siempre van acompañados de autovectores. Resolviendo tenemos que un autovector l.i. asociado a $\lambda = -1$ es $\mathbf{u}_1 = (1, -1)$.

para $\lambda = 6$,

$$(A - 6I)\mathbf{x} = \mathbf{0} \Rightarrow \begin{pmatrix} -5 & 2 \\ 5 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

De igual modo encontramos que un autovector l.i. asociado a $\lambda = 6$ es $\mathbf{u}_2 = (\frac{2}{5}, 1)$. Obsérvese que en virtud de (II) de la nota 6.1, no hay inconveniente en considerar un múltiplo de este vector, por ejemplo $(2, 5)$, para evitarnos las fracciones.

De este modo, la base $\mathcal{B}' = \{(1, -1), (2, 5)\}$ está formada por autovectores, y es posible diagonalizar $A$ del siguiente modo:

$$\begin{pmatrix} -1 & 0 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 5 \end{pmatrix}^{-1} A \begin{pmatrix} 1 & 2 \\ -1 & 5 \end{pmatrix}$$

(II) Sea $R_\theta$ la matriz de rotación en $\mathbb{R}^2$ que obtuvimos en el ejemplo 5.4:

$$R_\theta = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}$$

El polinomio característico es

$$|R_\theta - \lambda I| = \begin{vmatrix} \cos\theta - \lambda & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta - \lambda \end{vmatrix} = \lambda^2 - 2\lambda\cos\alpha + 1.$$

Sus raíces son $\lambda = \dfrac{2\cos\theta \pm \sqrt{-\operatorname{sen}^2\theta}}{2} = \cos\theta \pm i\operatorname{sen}\theta$.

Observamos que hay autovalores reales si y sólo si $\operatorname{sen}\theta = 0$, es decir si $\theta = k\pi$, $k \in \mathbb{Z}$. En este caso la matriz inicial $R_{k\pi}$ es ya diagonal, y obviamente diagonalizable.

Por su parte, cuando $\operatorname{sen}\theta \neq 0$, los autovalores están en $\mathbb{C}$ y son distintos, luego es diagonalizable (de forma compleja). Si embargo, si nos restringimos a $\mathbb{R}$, esta aplicación no sería diagonalizable pues no tiene autovalores reales, y por tanto no hay posibilidad de obtener autovectores reales que formen una base.

(III) Encontrar los valores del parámetro $a \in \mathbb{R}$ para los que la aplicación de matriz

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ a & 0 & 0 \end{pmatrix}$$

es diagonalizable.

El polinomio característico es

$$|A-\lambda I| = \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & a-\lambda & 0 \\ a & 0 & -\lambda \end{vmatrix} = \lambda^2(a-\lambda)$$

Los autovalores son $\lambda = 0$ (doble) y $\lambda = a$. Nótese que la matriz tendría un autovalor triple si $a = 0$, y un autovalor doble y otro simple si $a \neq 0$. Veamos cada uno de los casos:

Si $a = 0$, es decir $\lambda = 0$ es un autovalor triple, la matriz de partida ya es diagonal.

Si $a \neq 0$, $\lambda = a$ es un autovalor simple y $\lambda = 0$ un autovalor doble. Veamos si existe una base formada por autovectores. Para ello calcularemos los autovectores asociados a cada autovalor:

para $\lambda = a$,

$$\begin{pmatrix} -a & 0 & 0 \\ 0 & 0 & 0 \\ a & 0 & -a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \text{Autovector: } (0,1,0)$$

para $\lambda = 0$,

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ a & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \text{Autovector: } (0,0,1)$$

Puesto que $\{(0,1,0),(0,0,1)\}$ no forman una base de $\mathbb{R}^3$, la matriz no puede ser diagonalizable.[5]

En definitiva, para ver si una aplicación lineal es diagonalizable debemos calcular primero sus autovalores. Si estamos en un espacio de dimensión $n$, entonces habrá $n$ autovalores, contando sus multiplicidades, (véase la nota 6.3). Si los $n$ autovalores son distintos, el Teorema 6.2 nos dice que los $n$ autovectores

[5] Nótese que no puede haber más autovectores independientes, pues no tenemos más autovalores.

asociados son l.i. y por tanto tendremos una base formada por autovectores que nos asegura que la aplicación es diagonalizable.

Si por el contrario, alguno de los autovalores tiene multiplicidad mayor que uno, tendremos el número necesario de autovectores para formar una base siempre y cuando el número de autovectores asociados a cada autovalor múltiple coincida con la multiplicidad del mismo. Esto motiva la siguiente definición:

**Definición 6.5**

Se denomina *multiplicidad algebraica* de un autovalor a la multiplicidad que tiene como raíz del polinomio característico.

Se denomina *multiplicidad geométrica* de un autovalor al número de autovectores independientes asociados.

Por otro lado, el cálculo de autovectores que hemos llevado a cabo en el ejemplo anterior se puede realizar de forma más sistemática a través de los *subespacios propios*:

**Definición 6.6**

Sea $f : V \to V$ lineal con matriz asociada $A$ respecto de alguna base dada. Si $\lambda$ es un autovalor de $f$ se denomina *subespacio propio* correspondiente a $\lambda$ al subconjunto

$$E_1(\lambda) = \ker(A - \lambda I)$$

**Nota 6.4**

Obsérvese que $E_1(\lambda)$ es un subespacio invariante por $f$ (véase la (II) de la nota 6.1).

Usando el Teorema 5.3 y las definiciones 6.5 y 6.6 se tiene:

$$\dim(E_1(\lambda)) = \dim(V) - \text{rango}(A - \lambda I) = \text{multiplicidad geométrica}$$

Además se verifica el siguiente resultado que relaciona ambas multiplicidades:

**Proposición 6.2**

Para cada autovalor

$$1 \leq \text{multiplicidad geométrica} \leq \text{multiplicidad algebraica}$$

*Demostración:*
Sea $\mu$ un autovalor de multiplicidad algebraica $r$. La primera desigualdad es evidente pues el espacio $E_1(\mu)$ no puede ser el espacio nulo. Para ver la segunda desigualdad supongamos que la multiplicidad geométrica de $\mu$ es $h$, es decir, $E_1(\mu)$ es un espacio de dimensión $h$, con $\{\mathbf{u}_1, \dots, \mathbf{u}_h\}$ una base del mismo. Ampliemos la base anterior a una base del espacio total (que suponemos de dimensión $n$), y sea $A$ la matriz de la aplicación en dicha base. Por construcción, $A$ es de la forma

$$A = \left(\begin{array}{ccc|c} \mu & & & \\ & \ddots & & A' \\ & & \mu & \\ \hline & 0 & & A'' \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c}\mu\\ \ddots\\ \mu\end{array}}\right\} h \\ \}\, n-h \end{array}$$

pues los primeros $h$ vectores de la base son autovectores.

Si ahora calculamos el polinomio característico de $A$,

$$|A - \lambda I_n| = \left|\begin{array}{ccc|c} \mu-\lambda & & & \\ & \ddots & & A' \\ & & \mu-\lambda & \\ \hline & 0 & & A''-\lambda I_{n-h} \end{array}\right| = (\mu-\lambda)^h |A'' - \lambda I_{n-h}|$$

donde $|A'' - \lambda I_{n-h}|$ es un polinomio de grado $n-h$. Por tanto $\mu$ es una raíz de multiplicidad al menos $h$, es decir $r \geq h$.

■

Finalmente, damos una condición necesaria y suficiente para garantizar que una aplicación es diagonalizable en términos de las multiplicidades algebraica y geométrica de cada autovalor.

**Teorema 6.4**

Una aplicación lineal es diagonalizable si y solo si tiene $n$ autovalores, contando sus multiplicidades, y para cada autovalor, su multiplicidad algebraica coincide con su multiplicidad geométrica.

*Demostración:*
Supongamos que la aplicación $f$ es diagonalizable. En tal caso existirá una base $\mathcal{B}$ respecto de la cual la matriz de la aplicación será diagonal, es decir,

$$M_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Es claro que el polinomio característico de esta matriz es

$$P_f(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i)$$

luego $f$ tiene $n$ autovalores.
Si $\lambda_i$ es un autovalor simple, es evidente a partir de la Proposición 6.2 que su multiplicidad algebraica coincide con la geométrica. Si por el contrario $\lambda_j$ es un autovalor de multiplicidad $r$ (es decir, aparecerá $r$ veces en la diagonal de $M_{\mathcal{B}}(f)$, pongamos que en las columnas $j_1, \dots, j_r$) entonces el espacio $E_1(\lambda_j)$ tiene al menos $r$ vectores independientes, pues

$$f(\mathbf{u}_{j_1}) = \lambda_j \mathbf{u}_{j_1}, \ \dots, \ f(\mathbf{u}_{j_r}) = \lambda_j \mathbf{u}_{j_r}$$

con los $\mathbf{u}_{j_i}$ vectores de $\mathcal{B}$. Por otro lado, como la dimensión de ese espacio no puede sobrepasar la multiplicidad algebraica del autovalor (por la Proposición 6.2) se tendrá la igualdad entre multiplicidad algebraica y geométrica.

Recíprocamente, si para cada autovalor coincide su multiplicidad algebraica y geométrica significa que la suma directa de los espacios

$$E_1(\lambda_1) \oplus \cdots \oplus E_1(\lambda_k)$$

tiene dimensión $r_1 + \cdots + r_k = n$ (nótese que la intersección de cualesquiera de estos subespacios es el vector nulo en virtud del Teorema 6.2), luego dicha suma directa coincide con el espacio total. Es fácil probar que en tal situación la unión de las bases de estos subespacios es una base del espacio suma, y por tanto estará formada por autovectores. Por el Teorema 6.1, la aplicación es diagonalizable.

■

A partir de este resultado, identificar si una aplicación es o no diagonalizable resulta sencillo: basta calcular los autovalores y sus multiplicidades algebraicas (lo que se obtiene al resolver las raíces del polinomio característico) y geométrica (que consiste en estudiar la dimensión de los subespacios propios, que en última instancia no es más que el cálculo de rangos).

Analicemos en el siguiente ejemplo, qué sucede en el caso de matrices de orden dos.

**Ejemplo 6.4**

Consideremos $A \in \mathcal{M}_2(\mathbb{C})$ la matriz de una aplicación $f : V \to V$, con $\dim V = 2$. Veamos cuándo estas aplicaciones son diagonalizables.

El polinomio característico de $A$ será $\mathcal{P}_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)$, con $\lambda_1$ y $\lambda_2$ autovalores en $\mathbb{C}$. En tal caso se puede tener:

**$\lambda_1 \neq \lambda_2$.** Para esta situación ya hemos visto que los autovectores correspondientes a cada autovalor son l.i., por lo que tendremos una base formada por autovectores. El Teorema 6.1 asegura que la aplicación es diagonalizable. La matriz de la aplicación en la base de autovectores es

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

**$\lambda_1 = \lambda_2$.** Denotamos por $\lambda = \lambda_1 = \lambda_2$. Se dan dos situaciones:

(a) $\dim(\ker(A - \lambda I)) = 2$ (es decir, $\operatorname{rango}(A - \lambda I) = 0$). Esto significa que el número de autovectores l.i. asociados al autovalor existente es dos, por lo que tenemos una base formada por autovectores, y por tanto $A$ es diagonalizable. De hecho, si $\operatorname{rango}(A - \lambda I) = 0$ es que la matriz es diagonal, con $\lambda$ en la diagonal.

(b) $\dim(\ker(A - \lambda I)) = 1$ (es decir, $\operatorname{rango}(A - \lambda I) = 1$). En este caso, el número de autovectores independientes asociado al único autovalor es insuficiente para obtener una base de $V$, y por tanto $A$ no es diagonalizable.

Para matrices de orden tres, las posibilidades son más numerosas, por lo que veremos qué sucede en algunos ejemplos concretos.

**Ejemplo 6.5**

Estudiemos si la siguiente matriz es o no diagonalizable:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$$

Polinomio característico:

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 0 & 0 \\ 1 & -\lambda & -1 \\ 1 & -1 & -\lambda \end{vmatrix} = (1-\lambda)(\lambda^2 - 1)$$

Por tanto sus autovalores son: $\lambda = 1$ (doble), $\lambda = -1$ (simple).

Estudiemos los subespacios propios asociados a cada autovalor:

Para $\lambda = -1$,

$$E_1(-1) = \ker(A + I) \Rightarrow \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\dim(E_1(-1)) = 3 - \text{rango}(A + I) = 1$ y una base de $E_1(-1)$ está formada por el autovector $\mathbf{u}_1 = (0, 1, 1)$.

Para $\lambda = 1$,

$$E_1(1) = \ker(A - I) \Rightarrow \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\dim(E_1(1)) = 3 - \text{rango}(A - I) = 2$ y una base de $E_1(1)$ está formada por los autovectores $\mathbf{u}_2 = (1, 1, 0)$ y $\mathbf{u}_3 = (1, 0, 1)$.

Nótese que tenemos dos autovectores independientes asociados al autovalor 1 de multiplicidad dos, y un autovector asociado al autovalor $-1$ de multiplicidad uno (que es independiente con los otros dos debido al Teorema 6.2) luego tenemos una base de autovectores y por tanto la matriz $A$ es diagonalizable.

Si expresamos la matriz de esta aplicación en la base formada por los autovectores, esto es

$$\mathcal{B} = \{(0, 1, 1), (1, 1, 0), (1, 0, 1)\}$$

obtenemos la matriz

$$J_A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

de modo que $J_A = P^{-1}AP$, donde $P$ corresponde a la matriz de cambio de base de $\mathcal{B}$ a la base canónica, es decir

$$P = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

denominada comúnmente *matriz de paso*.

Por último nótese también que $E_1(1) \oplus E_1(-1) = V$.

**Ejemplo 6.6**

Estudiar si la siguiente matriz es o no diagonalizable:

$$A = \begin{pmatrix} 0 & 3 & 1 \\ 2 & -1 & -1 \\ -2 & -1 & -1 \end{pmatrix}$$

Comenzamos calculando el polinomio característico,

$$|A - \lambda I| = \begin{vmatrix} -\lambda & 3 & 1 \\ 2 & -1-\lambda & -1 \\ -2 & -1 & -1-\lambda \end{vmatrix} \overset{F_3+F_2}{=} \begin{vmatrix} -\lambda & 3 & 1 \\ 2 & -1-\lambda & -1 \\ 0 & -2-\lambda & -2-\lambda \end{vmatrix}$$

$$\overset{C_3-C_2}{=} \begin{vmatrix} -\lambda & 3 & -2 \\ 2 & -1-\lambda & \lambda \\ 0 & -2-\lambda & 0 \end{vmatrix} = (\lambda+2)(-\lambda^2+4)$$

Autovalores: $\lambda = 2$ (simple), $\lambda = -2$ (doble). Calculamos los subespacios propios asociados a cada autovalor (y con ello la multiplicidad geométrica).

Para $\lambda = 2$:

$$E_1(2) = \ker(A - 2I) \Rightarrow \begin{pmatrix} -2 & 3 & 1 \\ 2 & -3 & -1 \\ -2 & -1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\dim(E_1(2)) = 3 - \text{rango}(A-2I) = 1$, y resolviendo el sistema anterior obtenemos una base de $E_1(2)$ formada por el autovector $\mathbf{u}_1 = (-1,-1,1)$.

Para $\lambda = -2$:

$$E_1(-2) = \ker(A + 2I) \Rightarrow \begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Nuevamente $\dim(E_1(-2)) = 3 - \text{rango}(A + 2I) = 1$. Resolviendo el sistema anterior obtenemos una base de $E_1(-2)$ formada por el vector $(1,-1,1)$.

De aquí podemos deducir que la matriz $A$ no es diagonalizable, puesto que el número de autovectores independientes que proporciona este autovalor es insuficiente para formar una base junto con $\mathbf{u}_1$. Obsérvese que en este caso, $E_1(2) \oplus E_1(-2) \neq V$

6 3

## FORMA CANÓNICA DE JORDAN

En la sección anterior hemos visto las condiciones que aseguran que una aplicación es diagonalizable, pero como vimos en el ejemplo 6.2, tal proceso no siempre es posible. Puesto que el interés de la diagonalización reside en la posibilidad de trabajar con matrices de aplicaciones que sean lo más sencillas posibles (esto es, matrices diagonales), de forma natural podemos preguntarnos cuál es la matriz más sencilla asociada a una aplicación que no sea diagonalizable. Tal matriz se denomina *forma canónica de Jordan*.[6]

Comencemos viendo la situación en el caso de matrices de orden dos no diagonalizables. Recordemos que en el ejemplo 6.4, si los autovalores de una matriz $A \in \mathcal{M}_2(\mathbb{C})$ son iguales y $\text{rango}(A - \lambda I) = 1$ sabemos que la matriz no es diagonalizable. En este caso se tiene el siguiente resultado:

**Lema 6.3**

Si $A \in \mathcal{M}_2(\mathbb{K})$ y tiene sus dos autovalores iguales entonces la matriz $(A - \lambda I)^2 = 0$, donde $\lambda$ es el único autovalor de $A$.

*Demostración:*
Veamos que $(A - \lambda I)^2 \mathbf{x} = \mathbf{0}$, $\forall \mathbf{x} \in V$, por lo que $(A - \lambda I)^2$ será la aplicación nula, y por tanto $(A - \lambda I)^2 = \mathbf{0}$.

En primer lugar, si $\mathbf{x} \in E_1(\lambda)$ obviamente $(A - \lambda I)\mathbf{x} = \mathbf{0}$, de modo que trivialmente $(A - \lambda I)^2 \mathbf{x} = \mathbf{0}$. Supongamos entonces que $\mathbf{x} \notin E_1(\lambda)$, y sea $\mathbf{v}$ un autovector asociado a $\lambda$. Es evidente entonces que $\mathbf{x}$ y $\mathbf{v}$ conforman una base de $V$ (pues $V$ tiene dimensión dos, y ambos son independientes). Consideremos ahora

$$\mathbf{w} = (A - \lambda I)\mathbf{x} \in V \tag{6.3}$$

Como $\{\mathbf{x}, \mathbf{v}\}$ es una base podemos escribir $\mathbf{w} = \alpha \mathbf{x} + \beta \mathbf{v}$. Vamos a probar que $\alpha = 0$. Una vez probado esto, entonces ocurrirá que $\mathbf{w} \in E_1(\lambda)$ (pues es un múltiplo de $\mathbf{v}$), de modo que

$$\mathbf{0} = (A - \lambda I)\mathbf{w} = (A - \lambda I)^2 \mathbf{x}$$

como queríamos demostrar.

Veamos entonces que $\alpha = 0$, procediendo por reducción al absurdo. De (6.3) se tiene que $(A - \lambda I)\mathbf{x} = \alpha \mathbf{x} + \beta \mathbf{v}$, y operando llegamos a

$$(A - (\lambda + \alpha)I)\mathbf{x} = \beta \mathbf{v}$$

[6] Aparece en 1870 en el trabajo del matemático francés Camille Jordan titulado *Traité des substitutions et des équations algébriques.*

Suponiendo que $\alpha \neq 0$, como el único autovalor de $A$ es $\lambda$, entonces la matriz $B = A - (\lambda + \alpha)I$ tiene determinante no nulo, es decir, es no singular y por tanto podemos escribir

$$\mathbf{x} = B^{-1}\beta\mathbf{v} \tag{6.4}$$

Multiplicando por $\lambda$ esta última expresión, y teniendo en cuenta que los escalares conmutan con la multiplicación de matrices,

$$\lambda\mathbf{x} = \lambda\beta B^{-1}\mathbf{v} = \beta B^{-1}\lambda\mathbf{v} = \beta B^{-1}A\mathbf{v}$$

donde la última igualdad se debe a que $\mathbf{v}$ es un autovector de $A$ asociado a $\lambda$. Pero además, $A = B + (\lambda + \alpha)I$, de modo que lo anterior es:

$$\lambda\mathbf{x} = \beta B^{-1}(B + (\lambda + \alpha)I)\mathbf{v} = \beta\mathbf{v} + (\lambda + \alpha)\beta B^{-1}\mathbf{v}$$

Usando (6.4) se llega a que

$$\lambda\mathbf{x} = \beta\mathbf{v} + \lambda\mathbf{x} + \alpha\mathbf{x} \Rightarrow \beta\mathbf{v} + \alpha\mathbf{x} = \mathbf{0}$$

Esto es una combinación lineal de dos vectores que son independientes, igualados al vector nulo, es decir, $\alpha = \beta = 0$, que contradice nuestra suposición de que $\alpha \neq 0$. Así concluye la demostración.

■

Volvamos a la situación anterior. Denotemos por $E_1(\lambda) = \ker(A - \lambda I)$. En el caso que nos ocupa $E_1(\lambda)$ es un espacio de dimensión uno, y como estamos en un espacio $V$ de dimensión dos, podemos encontrar un vector no nulo $\mathbf{u}_2 \in V \backslash E$.

Sea ahora $\mathbf{u}_1 = (A - \lambda I)\mathbf{u}_2$. Entonces $\mathbf{u}_1 \in E_1(\lambda)$ pues

$$(A - \lambda I)\mathbf{u}_1 = (A - \lambda I)(A - \lambda I)\mathbf{u}_2 = \mathbf{0} \quad \text{(por el Lema 6.3)}.$$

Por tanto $\mathbf{u}_1$ es un autovector asociado al autovalor $\lambda$, luego $A\mathbf{u}_1 = \lambda\mathbf{u}_1$. Por otra parte,

$$\mathbf{u}_1 = (A - \lambda I)\mathbf{u}_2 \Rightarrow A\mathbf{u}_2 = \mathbf{u}_1 + \lambda\mathbf{u}_2.$$

Consideremos ahora la base $\{\mathbf{u}_1, \mathbf{u}_2\}$ (se trata de una base pues $\mathbf{u}_1 \in E_1(\lambda)$ y $\mathbf{u}_2 \in V \backslash E_1(\lambda)$). La matriz de la aplicación en esta base resulta:

$$J_A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

Esta es la forma canónica de Jordan de la matriz $A$.

**Ejemplo 6.7**

Encontrar la forma canónica de Jordan de

$$A = \begin{pmatrix} 0 & 2 \\ -2 & 4 \end{pmatrix}$$

El polinomio característico es

$$|A - \lambda I| = \begin{vmatrix} -\lambda & 2 \\ -2 & 4-\lambda \end{vmatrix} = (\lambda - 2)^2$$

con autovalores $\lambda = 2$ (doble).

Construimos $E_1(2) = \ker(A - 2I)$, que tiene dimensión 1, de manera que $A$ no puede ser diagonalizable. Para obtener la forma de Jordan calculamos una base de $E_1(2)$, que estará formada por un único vector, p.e. $(1, 1)$. Como este espacio resulta insuficiente para conseguir vectores de una base, consideramos el espacio

$$E_2(2) = \ker((A - 2I)^2)$$

Del Lema 6.3 sabemos que $(A - 2I)^2 = 0$, luego $E_2(2) = \mathbb{R}^2$, y entonces podemos escoger un vector de $\mathbb{R}^2 \backslash E_1(2)$, por ejemplo $\mathbf{u}_2 = (1, 0)$ (cualquier vector independiente con el anterior nos vale).

Ahora consideramos

$$\mathbf{u}_1 = (A - 2I)\mathbf{u}_2 \Rightarrow \begin{pmatrix} -2 & 2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

Nótese que $\mathbf{u}_1 \in E_1(2)$. Finalmente, consideramos la base $\{(-2, -2), (1, 0)\}$ (¡atención al orden!) y por tanto la aplicación $A$ tiene como matriz asociada en esta base

$$J_A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

La matriz de cambio de base será

$$P = \begin{pmatrix} -2 & 1 \\ -2 & 0 \end{pmatrix}$$

y se verifica $J_A = P^{-1}AP$.

**Ejemplo 6.8**

(I) (continuación del ejemplo 6.6)

Vimos que la matriz

$$A = \begin{pmatrix} 0 & 3 & 1 \\ 2 & -1 & -1 \\ -2 & -1 & -1 \end{pmatrix}$$

tiene autovalores $\lambda = 2$ (simple) y $\lambda = -2$ (doble) y subespacios propios $E_1(2) = L((-1,-1,1))$ y $E_1(-2) = L((1,-1,1))$

Sabíamos que la matriz $A$ no es diagonalizable porque el número de autovectores independientes que proporciona el autovalor doble es insuficiente para formar una base junto con el autovector asociado al autovalor simple.

También se observa que $E_1(2) + E_1(-2) \neq V$, es decir, existe un "agujero" entre los espacios de autovectores y el espacio total. Esta diferencia es debida a que el autovalor $\lambda = -2$ de multiplicidad algebraica 2 no genera un espacio de autovectores de suficiente dimensión (su multiplicidad geométrica es 1). Por ello consideramos ahora $E_2(-2) = \ker((A+2I)^2)$; para ello calculamos

$$(A+2I)^2 = \begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 8 & 0 \\ 8 & 8 & 0 \\ -8 & -8 & 0 \end{pmatrix}$$

Se tiene que $\dim E_2(-2) = 3 - \text{rango}((A+2I)^2) = 2$ y una base se obtendrá resolviendo el correspondiente sistema. En este caso,

$$\text{Base de } E_2(-2) = \{(-1,1,0),(0,0,1)\}$$

Es importante observar que $E_1(-2) \subset E_2(-2)$, como se comprueba fácilmente, ya que la base de $E_1(-2)$ está en $E_2(-2)$. Como el segundo espacio es más grande podemos tomar $\mathbf{u}_3 \in E_2(-2) \backslash E_1(-2)$. Por ejemplo, nosotros escogemos $\mathbf{u}_3 = (0,0,1)$.

A continuación calculamos $\mathbf{u}_2$ del siguiente modo:

$$\mathbf{u}_2 = (A+2I)\mathbf{u}_3 = \begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

Nótese ahora que $\mathbf{u}_2 \in E_1(-2)$, es decir es un autovector asociado a $\lambda = -2$ y por tanto satisface $A\mathbf{u}_2 = -2\mathbf{u}_2$. Por otra parte, está claro por construcción que $A\mathbf{u}_3 = \mathbf{u}_2 - 2\mathbf{u}_3$.

Por consiguiente, en la base $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, la aplicación dada tiene por matriz

$$J_A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -2 \end{pmatrix}$$

y la matriz de cambio de base es

$$P = \begin{pmatrix} -1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

(ii)

$$A = \begin{pmatrix} -2 & 1 & -1 \\ -1 & -1 & 0 \\ 0 & 1 & -3 \end{pmatrix}$$

Polinomio característico:

$$|A - \lambda I| = \begin{vmatrix} -2-\lambda & 1 & -1 \\ -1 & -1-\lambda & 0 \\ 0 & 1 & -3-\lambda \end{vmatrix} \overset{C_2+C_3}{=} \begin{vmatrix} -2-\lambda & 0 & -1 \\ -1 & -1-\lambda & 0 \\ 0 & -2-\lambda & -3-\lambda \end{vmatrix}$$

$$\overset{C_2-(1+\lambda)C_1}{=} \begin{vmatrix} -2-\lambda & (-2-\lambda)(-1-\lambda) & -1 \\ -1 & 0 & 0 \\ 0 & -2-\lambda & -3-\lambda \end{vmatrix}$$

$$= (-2-\lambda) \begin{vmatrix} -1-\lambda & -1 \\ 1 & -3-\lambda \end{vmatrix} = -(\lambda+2)^2$$

Autovalores $\lambda = -2$ (triple). Para calcular los autovectores asociados al único autovalor procedemos como antes:

$$E_1(-2) = \ker(A + 2I) \Rightarrow \begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\dim(E_1(-2)) = 3-\text{rango}(A+2I) = 1$ y una base de $E_1(-1)$ está formada por el autovector $(1,1,1)$. Nótese que como ya no hay más autovectores independientes con éste, la matriz $A$ no puede ser diagonalizable; una vez más, la multiplicidad algebraica y la geométrica no coinciden.

Procedemos como antes calculando el espacio $E_2(-2) = \ker((A+2I)^2)$:

$$(A+2I)^2 = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

$\dim(E_2(-2)) = 3 - \text{rango}((A+2I)^2) = 2$ y una base de $E_2(-2)$ obtenida al resolver el sistema con matriz anterior es la formada por $(0,1,0)$ y $(1,0,1)$. Obsérvese que $E_1(-2) \subset E_2(-2)$. Al igual que antes podremos coger un vector de $E_2(-2)$ que no esté en $E_1(-2)$, pero uno solo, pues la dimensión de $E_2(-2)$ no permite más. Lo que hacemos en este caso es seguir ampliando el espacio con $E_3(-2) = \ker((A+2I)^3)$. Un simple cálculo nos muestra que $(A+2I)^3 = \mathbf{0}$. Es decir, $E_3(-2) = V$. Tenemos por tanto la inclusión de espacios siguiente

$$E_1(-2) \subset E_2(-2) \subset E_3(-2) = V$$

Para construir la forma canónica de Jordan y la matriz de paso correspondiente consideramos un vector del espacio mayor que no pertenezca al espacio inmediatamente inferior, esto es, $\mathbf{u}_3 \in E_3(-2)\backslash E_2(-2)$, por ejemplo, $\mathbf{u}_3 = (1,0,0)$. Sean ahora $\mathbf{u}_2 = (A+2I)\mathbf{u}_3$ y $\mathbf{u}_1 = (A+2I)\mathbf{u}_2$, es decir,

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

Por construcción, el vector $\mathbf{u}_2 \in E_2(-2)$ mientras que $\mathbf{u}_1 \in E_1(-2)$, es decir, es un autovector. Luego

$$A\mathbf{u}_1 = -2\mathbf{u}_1, \quad A\mathbf{u}_2 = \mathbf{u}_1 - 2\mathbf{u}_2, \quad A\mathbf{u}_3 = \mathbf{u}_2 - 2\mathbf{u}_3$$

Si escribimos la aplicación en la base $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ obtenemos la forma de

Jordan y la matriz de paso siguientes:

$$J_A = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -2 \end{pmatrix}, \quad P = \begin{pmatrix} -1 & 0 & 1 \\ -1 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

(III)

$$A = \begin{pmatrix} -2 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

Polinomio característico:

$$|A - \lambda I| = \begin{vmatrix} -2-\lambda & 0 & 1 \\ 0 & -1-\lambda & 0 \\ -1 & 0 & -\lambda \end{vmatrix} = (-1-\lambda)(\lambda(2+\lambda)+1) = -(\lambda+1)^3$$

Autovalores $\lambda = -1$ (triple). Cálculo de autovectores:

$$E_1(-1) = \ker(A+I) \Rightarrow \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\dim(E_1(-2)) = 3 - \text{rango}(A+I) = 2$ y una base de $E_1(-1)$ está formada por los autovectores $(0,1,0)$ y $(1,0,1)$. Nuevamente el número de autovectores independientes es insuficiente para formar una base y por tanto la matriz $A$ no es diagonalizable.

Como necesitamos más vectores, los buscamos en el espacio $E_2(-1)$:

$$(A+I)^2 = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

luego $E_2(-1) = V$. Tenemos ahora la cadena de subespacios

$$E_1(-2) \subset E_2(-2) = V$$

Tomamos un vector $\mathbf{u}_3 \in E_2(-1) \setminus E_1(-1)$, por ejemplo $\mathbf{u}_3 = (1,0,0)$, y calculamos $\mathbf{u}_2 = (A+I)\mathbf{u}_3$:

$$\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} \in E_1(-1)$$

luego $\mathbf{u}_2$ es un autovector. Como en el espacio $E_1(-1)$ tenemos dos autovectores independientes tomamos $\mathbf{u}_1 \in E_1(-1)$, independiente con este último, por ejemplo, $\mathbf{u}_1 = (0, 1, 0)$.

La forma de Jordan y la correspondiente matriz de paso son:

$$J_A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \quad P = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

En los ejemplos anteriores hemos visto cómo tratar el caso de matrices no diagonalizables de dimensión dos y tres. A continuación vamos a sistematizar el método para matrices de cualquier orden.

**Definición 6.7**

Dos matrices $A$ y $B$ se dicen *equivalentes* si existe una matriz invertible $P$ tal que $B = P^{-1}AP$.

En particular, las matrices de un endomorfismo en diferentes bases son equivalentes.

**Definición 6.8**

Se denomina *matriz elemental de Jordan de orden $k$ (o caja de orden $k$) y autovalor* $\lambda \in \mathbb{C}$ a la matriz $J_k(\lambda) \in \mathcal{M}_k(\mathbb{C})$ con elementos nulos excepto en la diagonal (donde aparece $\lambda$) y encima de la misma (en la que aparecen 1).

Algunas cajas elementales de Jordan son

$$J_1(\lambda) = (\lambda), \quad J_2(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad J_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

**Definición 6.9**

Se denomina *matriz de Jordan* a cualquier matriz cuadrada formada por yuxtaposición de matrices elementales de Jordan a lo largo de su diagonal, es decir, una matriz *diagonal por cajas* en la que cada caja es una matriz elemental.

$$J = \begin{pmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_n}(\lambda_n) \end{pmatrix}$$

Por ejemplo, las siguientes son matrices de Jordan formadas "pegando" cajas de diversos tamaños y/o autovalores:

$$\left(\begin{array}{cc|c|c} 3 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ \hline 0 & 0 & 3 & 0 \\ \hline 0 & 0 & 0 & 2 \end{array}\right) \quad \left(\begin{array}{cccc|c} 5 & 1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right) \quad \left(\begin{array}{c|ccc|c} 5 & 0 & 0 & 0 & 0 \\ \hline 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 1 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

Sin embargo, las siguientes matrices no son matrices de Jordan; ¿puede el lector averiguar la razón?

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 5 & 1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix} \quad \begin{pmatrix} 5 & 1 & 0 & 0 & 1 \\ 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

**Teorema 6.5 (Teorema de Jordan (forma compleja))**

Toda matriz $A \in \mathcal{M}_n(\mathbb{C})$ es equivalente a una matriz de Jordan $J \in \mathcal{M}_n(\mathbb{C})$ que es única, salvo permutación de las cajas que la forman.

La prueba del Teorema de Jordan está basada en el siguiente resultado sobre los subespacios $E_j(\lambda) = \ker((A - \lambda I)^j)$, denominados *subespacios propios generalizados*.

**Proposición 6.4**

Sea $\lambda$ un autovalor de $A$ de multiplicidad $r$ y consideremos los subespacios $E_j(\lambda)$, $j \geq 0$. Entonces:

(I) $E_j(\lambda)$ es una sucesión estrictamente creciente de subespacios de manera que existe $t$ tal que $E_t(\lambda) = E_{t+k}(\lambda)$, $\forall k \geq 0$; esto es,

$$\{\mathbf{0}\} = E_0(\lambda) \subsetneq E_1(\lambda) \subsetneq E_2(\lambda) \subsetneq \cdots \subsetneq E_t(\lambda) = E_{t+1}(\lambda) = \cdots$$

A $E_t(\lambda)$ se le denomina *subespacio máximo* asociado al autovalor $\lambda$

(II) $E_t(\lambda)$ es un subespacio invariante.

(III) $\mathbf{u} \in E_j(\lambda)$ si y solo si $(A - \lambda I)\mathbf{u} \in E_{j-1}(\lambda)$.

(IV) Se verifica que $\dim(E_t(\lambda)) = r$

Además, si $\lambda_1, \ldots, \lambda_k$ son los autovalores de $A$ de multiplicidades respectivas $r_1, \ldots, r_k$, con $n = r_1 + \cdots + r_k$, y $E_{t_j}(\lambda_j)$, $j = 1, \ldots, k$ son los subespacios máximos asociados a cada autovalor, entonces

$$V = E_{t_1}(\lambda_1) \oplus \cdots \oplus E_{t_k}(\lambda_k) \tag{6.5}$$

donde $V$ es el espacio ambiente.

*Demostración:*
La demostración completa de este resultado precisa herramientas que no hemos contemplado en este texto, por lo que solo probaremos (I)–(III).

(I) Que la sucesión de espacios $E_j(\lambda)$ es creciente es sencillo de probar, pues si $\mathbf{v} \in E_j(\lambda)$, eso significa que

$$(A - \lambda I)^j \mathbf{v} = \mathbf{0} \Rightarrow (A - \lambda I)(A - \lambda I)^j \mathbf{v} = (A - \lambda I)\mathbf{0} = \mathbf{0}$$

es decir, $(A - \lambda I)^{j+1}\mathbf{v} = \mathbf{0}$, lo que significa que $\mathbf{v} \in E_{j+1}(\lambda)$.

Por otra parte, si $E_t(\lambda) = E_{t+1}(\lambda)$, entonces es fácil ver que $E_t(\lambda) = E_{t+k}(\lambda)$, $\forall k \geq 0$. En efecto, si $\mathbf{v} \in E_{t+2}$, entonces

$$\mathbf{0} = (A - \lambda I)^{t+2}\mathbf{v} = (A - \lambda I)^{t+1}(A - \lambda I)\mathbf{v}$$

de modo que $(A - \lambda I)\mathbf{v} \in E_{t+1}(\lambda) = E_t(\lambda)$, lo que significa que

$$(A - \lambda I)^t((A - \lambda I)\mathbf{v}) = \mathbf{0} \Rightarrow (A - \lambda I)^{t+1}\mathbf{v} = \mathbf{0}$$

de donde concluimos que $\mathbf{v} = E_{t+1}(\lambda)$, y de aquí $E_{t+2}(\lambda) \subset E_{t+1}(\lambda)$. La inclusión contraria ya la hemos probado al ver que los subespacios son

crecientes; es decir, $E_{t+1}(\lambda) = E_{t+2}(\lambda)$. El mismo razonamiento se usa para probar las igualdades siguientes.

Finalmente, debe existir $t$ tal que $E_t(\lambda) = E_{t+1}(\lambda)$ pues tenemos una sucesión creciente de subespacios vectoriales dentro de un espacio de dimensión finita, por lo que no puede crecer indefinidamente.

Como consecuencia nótese que las dimensiones de los subespacios propios generalizados conforman una sucesión estrictamente creciente hasta que se vuelve constante una vez llegado al subespacio máximo.

(ii) Veamos que si $\mathbf{u} \in E_t(\lambda)$ entonces $A\mathbf{u}$ también pertenece a dicho conjunto. Esto es debido a

$$\begin{aligned}(A-\lambda I)^t(A\mathbf{u}) &= (A-\lambda I)^t(A\mathbf{u}-\lambda\mathbf{u}+\lambda\mathbf{u}) = (A-\lambda I)^t((A-\lambda I)\mathbf{u}+\lambda\mathbf{u}) \\ &= (A-\lambda I)^{t+1}\mathbf{u} + \lambda(A-\lambda I)^t\mathbf{u} = \mathbf{0} + \lambda\mathbf{0} = \mathbf{0}\end{aligned}$$

(iii) Está claro que $\mathbf{u} \in E_j(\lambda)$ si y solo si

$$\mathbf{0} = (A-\lambda I)^j\mathbf{u} = (A-\lambda I)^{j-1}((A-\lambda I)\mathbf{u})$$

es decir, si $(A-\lambda I)\mathbf{u} \in E_{j-1}(\lambda)$.

■

### Construcción de la forma compleja de Jordan y de una matriz de paso

Dada una matriz $A \in \mathcal{M}_n(\mathbb{C})$, para construir la matriz de Jordan seguiremos los siguientes pasos:

(a) Calcular los autovalores de $A$, $\lambda_1, \dots, \lambda_k$ junto con sus multiplicidades algebraicas $r_1, \dots, r_k$. Puesto que estamos en $\mathbb{C}$ sabemos que $r_1 + \cdots + r_k = n$.

(b) Para cada autovalor $\lambda$ de multiplicidad $r$ construimos los espacios $E_j(\lambda) = \ker((A-\lambda I)^j)$, con $j \geq 0$ hasta llegar al subespacio máximo, es decir, aquél que verifica que $\dim(E_t(\lambda)) = r$. Denotaremos por

$$q_j = \dim(E_j(\lambda)), \quad j = 0, \dots, t \tag{6.6}$$

(c) Construimos la denominada *partición de multiplicidad* del autovalor $\lambda$. Para ello, dados los $q_j$ calculados en (6.6), definimos

$$p_j = q_j - q_{j-1}, \quad j = 1, \dots, t$$

Obsérvese que $q_j$ es una sucesión estrictamente creciente de números naturales, mientras que $p_j$ es una sucesión decreciente. Se define la **partición de multiplicidad** del autovalor $\lambda$ como la expresión

$$r = p_1 + \cdots + p_t$$

(d) Construida la partición de multiplicidad, la forma canónica de Jordan tiene, para cada autovalor:

$$\left\{\begin{array}{ccc} p_t & \text{cajas de tamaño} & t \\ p_{t-1}-p_t & \text{''} & t-1 \\ p_{t-2}-p_{t-1} & \text{''} & t-2 \\ \vdots & \vdots & \vdots \\ p_1-p_2 & \text{''} & 1 \end{array}\right. \tag{6.7}$$

**Nota 6.5**

La forma canónica de Jordan sólo depende de la partición de multiplicidad de cada uno de sus autovalores.

Para construir la matriz de paso procedemos como sigue:

(a) Consideramos (a nuestra elección) $p_t$ vectores de $E_t(\lambda)\backslash E_{t-1}(\lambda)$ linealmente independientes entre sí, que denotaremos por $\mathbf{u}_1^t,\dots,\mathbf{u}_{p_t}^t$

(b) Para cada uno de los vectores anteriores construimos

$$\mathbf{u}_l^{t-1}=(A-\lambda I)\mathbf{u}_l^t, \quad l=1,\dots,p_t \tag{6.8}$$

y los completamos con vectores de $E_{t-1}(\lambda)\backslash E_{t-2}(\lambda)$ linealmente independientes hasta obtener un total de $p_{t-1}$ vectores. Ahora en este nivel tenemos

$$\{\mathbf{u}_1^t,\dots,\mathbf{u}_{p_t}^t\}\subset E_t(\lambda),\ \{\mathbf{u}_1^{t-1},\dots,\mathbf{u}_{p_t}^{t-1},\mathbf{u}_{p_t+1}^{t-1},\dots,\mathbf{u}_{p_{t-1}-p_t}^{t-1}\}\subset E_{t-1}(\lambda)$$

(c) Repetimos la operación del paso anterior sobre el conjunto de vectores de $E_{t-1}(\lambda)$, es decir, aplicamos $(A-\lambda I)$ a cada uno de los vectores anteriores y los completamos con vectores independientes de $E_{t-2}\backslash E_{t-3}$ hasta obtener $p_{t-2}$ vectores.

Podemos usar la siguiente representación gráfica de la construcción de la base:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $p_t$ | $E_t(\lambda)$ | $\mathbf{u}_1^t$ | $\cdots$ | $\mathbf{u}_{p_t}^t$ | | | | |
| $p_{t-1}$ | $E_{t-1}(\lambda)$ | $\mathbf{u}_1^{t-1}$ | $\cdots$ | $\mathbf{u}_{p_t}^{t-1}$ | $\mathbf{u}_{p_t+1}^{t-1}$ | $\cdots$ | $\mathbf{u}_{p_{t-1}-p_t}^{t-1}$ | |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\dots$ |
| $p_1$ | $E_1(\lambda)$ | $\mathbf{u}_1^1$ | $\cdots$ | $\mathbf{u}_{p_t}^1$ | $\mathbf{u}_{p_t+1}^1$ | $\cdots$ | $\mathbf{u}_{p_{t-1}-p_t}^1$ | $\cdots$ |

(6.9)

**En cada una de las filas correspondiente a un subespacio $E_j(\lambda)$ debemos tener exactamente $p_j$ vectores linealmente independientes**. Conocidos los vectores del espacio $E_j(\lambda)$, construimos los vectores de $E_{j-1}(\lambda)$ multiplicando cada uno de ellos por $(A - \lambda I)$. Si el número de vectores así obtenido es inferior a $p_{j-1}$, añadimos vectores linealmente independientes de $E_{j-1}(\lambda)\backslash E_{j-2}(\lambda)$ en número necesario hasta obtener $p_{j-1}$ vectores.

Finalmente, los vectores que formarán parte de la matriz de paso son los vectores de cada uno de los subespacios anteriores escritos **por columnas de abajo hacia arriba** (¡atención al orden![7]), es decir

$$\{\mathbf{u}_1^1, \mathbf{u}_1^2, \dots, \mathbf{u}_1^t, \mathbf{u}_2^1, \dots, \mathbf{u}_2^t, \dots, \mathbf{u}_{p_t}^1, \dots, \mathbf{u}_{p_t}^t, \mathbf{u}_{p_t+1}^1, \dots \mathbf{u}_{p_t+1}^{t-1}, \mathbf{u}_{p_t+2}^1, \dots\} \tag{6.10}$$

**Nota 6.6**

La tabla anterior también nos proporciona de forma gráfica el número y el tamaño de las cajas elementales de Jordan que aporta cada autovalor. Hemos de considerar tantas cajas como columnas de vectores tenemos en la tabla, y cada caja tendrá un tamaño igual al número de vectores que hay en cada columna.

*Demostración:* (Teorema de Jordan)
Para demostrar finalmente el resultado central probaremos dos cosas:

(i) Si $\lambda$ es un autovalor del endomorfismo $A$, los vectores obtenidos en (6.10) forman una base de $E_t(\lambda)$, siendo $E_t(\lambda)$ el subespacio máximo asociado a $\lambda$. Además, la restricción de $A$ al subespacio $E_t(\lambda)$ tiene como matriz asociada una matriz de Jordan, cuyas cajas corresponden a (6.7).

(ii) Si $V = M_1 \oplus \cdots \oplus M_r$, con $M_i$ subespacios invariantes por $A$, y $A_i$ son las matrices de la aplicación $A$ restringida a cada uno de estos subespacios, entonces existe una base de $V$ tal que la matriz de $A$ es diagonal por cajas, en el que cada caja corresponde a una de la matrices $A_i$.

A partir de estos dos hechos y usando (6.5) el resultado es inmediato.

Veamos (i): para probar que los vectores de (6.10) forman base solo hay que tener en cuenta su construcción y (iii) de la Proposición 6.4, puesto que los vectores elegidos en $E_j(\lambda)\backslash E_{j-1}(\lambda)$ son independientes y los vectores construidos en $E_{j-1}(\lambda)$ a partir de la operación (6.8) también serán independientes.

[7] El orden de estos vectores debe venir de acuerdo al orden impuesto sobre las cajas elementales en la forma de Jordan. El orden descrito aquí corresponde a una ordenación de las cajas de mayor a menor tamaño.

Como tenemos tantos vectores como dimensión de $E_t(\lambda)$, forman una base del subespacio máximo.

Por otra parte, si restringimos la aplicación $A$ al subespacio $E_t(\lambda)$ y calculamos la matriz de esta aplicación en la base formada por los vectores de (6.10), debido a (6.8), vemos que si $t > 1$,

$$A\mathbf{u}_l^t = \lambda\mathbf{u}_l^t + \mathbf{u}_l^{t-1}$$

de modo que las coordenadas en esta base son $(0, \dots, 0, 1, \lambda, 0, \dots, 0)$, mientras que si $t = 1$, los vectores $\mathbf{u}_j^1 \in E_1(\lambda)$, luego son autovectores y por tanto $A\mathbf{u}_j^1 = \lambda\mathbf{u}_j^1$, es decir, las coordenadas son $(0, \dots, 0, \lambda, 0, \dots, 0)$. En definitiva, por cada columna de vectores de la tabla (6.9) tenemos una caja elemental de Jordan de tamaño igual al número de vectores existente en cada columna.

Veamos (ii): para simplificar los detalles, supongamos que $V = M_1 \oplus M_2$, con $M_1$ y $M_2$ subespacios invariantes por $A$. Debido a que $V$ es la suma directa podemos encontrar una base de $V$ juntando las bases de $M_1$ y $M_2$. Pongamos que

$$\text{Base de } M_1 = \{\mathbf{e}_1, \dots, \mathbf{e}_k\}, \quad \text{Base de } M_2 = \{\mathbf{e}_{k+1}, \dots, \mathbf{e}_n\}$$

Como ambos espacios son invariantes, está claro que

$$A\mathbf{e}_j = \sum_{i=1}^{k} \alpha_{i,j}\mathbf{e}_i, \quad j = 1, \dots, k$$

mientras que

$$A\mathbf{e}_j = \sum_{i=k+1}^{n} \beta_{i,j}\mathbf{e}_i, \quad j = k+1, \dots, n$$

Si $A_1 = (\alpha_{i,j}) \in \mathcal{M}_k(\mathbb{C})$ y $A_2 = (\beta_{i,j}) \in \mathcal{M}_{n-k+1}(\mathbb{C})$ entonces es fácil ver que la matriz de $A$ en la base $\{\mathbf{e}_1, \dots \mathbf{e}_n\}$ es

$$A = \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix}$$

La extensión a $r$ subespacios invariantes es inmediata.

■

Veamos el funcionamiento del método de construcción de la forma de Jordan sobre algunos ejemplos.

**Ejemplo 6.9**

$$A = \begin{pmatrix} 1 & 0 & 2 & -6 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Comenzamos calculando el polinomio característico:

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 0 & 2 & -6 \\ 0 & 1-\lambda & -1 & 3 \\ 0 & 0 & 1-\lambda & 3 \\ 0 & 0 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)^3(2-\lambda)$$

Autovalores: $\lambda = 1$ (triple) y $\lambda = 2$ (simple).

Comenzamos el estudio para el autovalor $\lambda = 1$ de multiplicidad $r = 3$. Construimos el espacio de autovectores asociado $E_1(1) = \ker(A - I)$:

$$A - I = \begin{pmatrix} 0 & 0 & 2 & -6 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \text{rango}(A - I) = 2$$

Luego $q_1 = \dim E_1(1) = 4 - 2 = 2$. Para obtener una base de este espacio resolvemos el sistema (formado por las filas segunda y cuarta, con las que se tiene rango dos)

$$\left.\begin{aligned} -x_3 + 3x_4 &= 0 \\ x_4 &= 0 \end{aligned}\right\} \Rightarrow \text{Base de } E_1(1) = \begin{cases} (1,0,0,0) \\ (0,1,0,0) \end{cases}$$

Puesto que el número de autovectores independientes asociado a este autovalor no coincide con la multiplicidad del mismo (lo cual ya significa que la matriz no es diagonalizable), procedemos con el siguiente espacio $E_2(1)$:

$$(A - I)^2 = \begin{pmatrix} 0 & 0 & 2 & -6 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 2 & -6 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

cuyo rango es uno. Por tanto $q_2 = \dim E_2(1) = 3$, y puesto que es igual a la multiplicidad del autovalor, éste es el subespacio máximo. Obsérvese que sin

necesidad de calcular la matriz $(A - I)^2$ podríamos haber sabido de antemano la dimensión de este subespacio: téngase en cuenta que los subespacios propios generalizados son estrictamente crecientes hasta que se estabilizan, justo cuando alcanzan la dimensión máxima, que siempre es igual a la multiplicidad del autovalor.

Es fácil encontrar una base de $E_2(1)$ resolviendo el correspondiente sistema; en este caso obtenemos:

$$\text{Base de } E_2(1) = \{(1,0,0,0),(0,1,0,0),(0,0,1,0)\}$$

Así pues, tenemos que $q_0 = 0$, $q_1 = 2$ y $q_2 = 3$ y por tanto,

$$p_1 = q_1 - q_0 = 2, \quad p_2 = q_2 - q_1 = 1$$

La partición de multiplicidades se escribe $r = p_1 + p_2$, es decir $3 = 2 + 1$.

A esta partición le corresponden:

$$p_2 = 1 \text{ caja de tamaño } 2$$
$$p_1 - p_2 = 1 \text{ caja de tamaño } 1$$

Así pues, la aportación del autovalor $\lambda = 1$ a la forma de Jordan es

$$\left(\begin{array}{cc|c} 1 & 1 & \\ 0 & 1 & \\ \hline & & 1 \end{array}\right)$$

Para construir los vectores correspondientes a este autovalor que aparecerán en la matriz de paso construimos el esquema

| | | | |
|---|---|---|---|
| $p_2 = 1$ | $E_2(1)$ | $\mathbf{u}_1^2 = (0,0,1,0)$ | |
| $p_1 = 2$ | $E_1(1)$ | $\mathbf{u}_1^1 = (2,-1,0,0)$ | $\mathbf{u}_2^1 = (1,0,0,0)$ |

donde $\mathbf{u}_1^2$ ha sido escogido arbitrariamente en $E_2(1)\backslash E_1(1)$, $\mathbf{u}_1^1$ se ha calculado mediante $\mathbf{u}_1^1 = (A - I)\mathbf{u}_1^2$, mientras que $\mathbf{u}_2^1$ se ha tomado arbitrariamente en $E_1(1)$, linealmente independiente con $\mathbf{u}_1^1$. Vemos que la tabla nos indica también el número de cajas: hay un columna formada por dos vectores y una columna formada por uno (esto es, una caja de tamaño dos, y una caja de tamaño uno).

La aportación de este autovalor a la matriz de paso es (insistimos en el orden)

$$(2,-1,0,0),(0,0,1,0),(1,0,0,0)$$

Con respecto al autovalor $\lambda = 2$, puesto que su multiplicidad es uno, el subespacio máximo asociado no puede ser más que $E_1(2) = \ker(A - 2I)$. En

este caso,

$$A - 2I = \begin{pmatrix} -1 & 0 & 2 & -6 \\ 0 & -1 & -1 & 3 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Sin necesidad de calcularlo, el rango$(A - 2I) = 3$ (puesto que la dimensión de este subespacio debe ser uno). Una base se obtendrá resolviendo el sistema correspondiente,

$$\left.\begin{array}{rcl} -x_1 + 2x_3 - 6x_4 & = & 0 \\ -x_2 - x_3 + 3x_4 & = & 0 \\ x_3 + 3x_4 & = & 0 \end{array}\right\} \Rightarrow \text{Base de } E_1(2) = \{(0,0,3,1)\}$$

En el caso de autovalores simples no hay necesidad de efectuar la partición de multiplicidad puesto que solo le puede corresponder una caja elemental de tamaño uno.

Como conclusión, la forma de Jordan de $A$ se escribe yuxtaponiendo las cajas obtenidas para cada autovalor,

$$J_A = \left(\begin{array}{cc|c|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 2 \end{array}\right)$$

y la matriz de paso escribiendo, por columnas, los vectores obtenidos:

$$P = \begin{pmatrix} 2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

verificándose que $J_A = P^{-1}AP$.

Veamos un nuevo ejemplo en el que van a intervenir valores complejos.

**Ejemplo 6.10**

Calcular la forma de Jordan de la matriz

$$A = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & -2 & 0 & 1 \end{pmatrix}$$

Polinomio característico:

$$|A-\lambda I| = \begin{vmatrix} -\lambda & 1 & -1 & 0 \\ 0 & -1-\lambda & 0 & 1 \\ 1 & 1 & -\lambda & 0 \\ 0 & -2 & 0 & 1-\lambda \end{vmatrix}$$

$$\overset{F_1+\lambda F_3}{=} \begin{vmatrix} 0 & \lambda+1 & -1-\lambda^2 & 0 \\ 0 & -1-\lambda & 0 & 1 \\ 1 & 1 & -\lambda & 0 \\ 0 & -2 & 0 & 1-\lambda \end{vmatrix} = (-1-\lambda^2) \begin{vmatrix} -1-\lambda & 1 \\ -2 & 1-\lambda \end{vmatrix} = (\lambda^2+1)^2$$

Autovalores: $\lambda = i$ (doble) y $\lambda = -i$ (doble). Nótese que para poder aplicar el Teorema de Jordan debemos usar raíces complejas.

Para el autovalor $\lambda = i$ de multiplicidad $r = 2$ consideramos el espacio de autovectores $E_1(i) = \ker(A - iI)$, donde

$$A - iI = \begin{pmatrix} -i & 1 & -1 & 0 \\ 0 & -1-i & 0 & 1 \\ 1 & 1 & -i & 0 \\ 0 & -2 & 0 & 1-i \end{pmatrix} \Rightarrow \text{rango}(A-iI) = 3$$

ya que el menor

$$\begin{vmatrix} -i & 1 & 0 \\ 0 & -1-i & 1 \\ 1 & 1 & 0 \end{vmatrix} \neq 0$$

(nótese que no es necesario calcular $|A - iI|$, pues sabemos que debe ser nulo).

Entonces $q_1 = \dim E_1(i) = 4 - 3 = 1$ y obtenemos una base de este espacio al resolver el sistema correspondiente al menor no nulo anterior:

$$\left.\begin{aligned} -ix_1 + x_2 - x_3 &= 0 \\ (-1-i)x_2 + x_4 &= 0 \\ x_1 + x_2 - ix_3 &= 0 \end{aligned}\right\} \Rightarrow \text{para } x_3 = 1 \text{ resolvemos } \left.\begin{aligned} -ix_1 + x_2 &= 1 \\ -x_1 - x_2 &= -i \end{aligned}\right\}$$

de donde $x_2 = 0$, $(-1-i)x_1 = 1-i \Rightarrow x_1 = i$. Finalmente se tiene que $x_4 = 0$. Una base de $E_1(i)$ está formada por el autovector $(i, 0, 1, 0)$. Como el número de autovectores independientes asociado a este autovalor no coincide con la multiplicidad del mismo (es decir, la matriz no es diagonalizable), calculamos el espacio $E_2(i)$,

$$(A - iI)^2 = \begin{pmatrix} -i & 1 & -1 & 0 \\ 0 & -1-i & 0 & 1 \\ 1 & 1 & -i & 0 \\ 0 & -2 & 0 & 1-i \end{pmatrix} \begin{pmatrix} -i & 1 & -1 & 0 \\ 0 & -1-i & 0 & 1 \\ 1 & 1 & -i & 0 \\ 0 & -2 & 0 & 1-i \end{pmatrix}$$

$$= \begin{pmatrix} -2 & -2-2i & 2i & 1 \\ 0 & -2+2i & 0 & -2i \\ -2i & -2i & -2 & 1 \\ 0 & 4i & 0 & -2-2i \end{pmatrix}$$

El rango$((A-iI)^2) = 2$ y por tanto $q_2 = \dim E_2(1) = 2$. Puesto que la dimensión de este subespacio coincide con la multiplicidad del autovalor, estamos ante el subespacio máximo. Resolviendo el sistema

$$\left.\begin{array}{rcl} -2x_1 + (-2-2i)x_2 + 2ix_3 + x_4 & = & 0 \\ (-2-2i)x_2 - 2ix_4 & = & 0 \end{array}\right\}$$

con parámetros $x_3$ y $x_4$ obtenemos los vectores $(i, 0, 1, 0), (-\frac{1}{2}, \frac{1}{2} - \frac{1}{2}i, 0, 1)$.

Nótese que podemos reemplazar el vector $(-\frac{1}{2}, \frac{1}{2} - \frac{1}{2}i, 0, 1)$ por un múltiplo suyo, ya que seguiríamos teniendo una base de $E_2(i)$, así pues, por comodidad, escogemos como base $\{(i, 0, 1, 0), (-1, 1-i, 0, 2)\}$.

Ya tenemos la información necesaria para encontrar las cajas correspondientes a este autovalor y los vectores que aporta a la matriz de paso. Comenzamos con la partición de multiplicidades:

$$q_0 = 0,\ q_1 = 1,\ q_2 = 2 \Rightarrow p_1 = 1,\ p_2 = 1$$

Es decir, $2 = 1 + 1$. A esta partición le corresponde una caja de tamaño dos y cero cajas de tamaño uno, es decir,

$$\begin{pmatrix} i & 1 \\ 0 & i \end{pmatrix}$$

Los vectores aportados a la matriz de paso se obtienen según la tabla siguiente:

| | | |
|---|---|---|
| $p_2 = 1$ | $E_2(i)$ | $\mathbf{u}_1^2 = (-1, 1-i, 0, 2)$ |
| $p_1 = 1$ | $E_1(i)$ | $\mathbf{u}_1^1 = (1, 0, -i, 0)$ |

donde $\mathbf{u}_1^2$ ha sido escogido en $E_2(i)\backslash E_1(i)$, mientras que $\mathbf{u}_1^1 = (A - iI)\mathbf{u}_1^2$.

Como se ha podido comprobar, el cálculo de subespacios propios cuando el autovalor es un número complejo es ciertamente tedioso. Sin embargo, como vamos a ver a continuación, en los **casos de matrices reales con autovalores complejos** como el que nos ocupa, se presenta cierta "simetría" que facilita mucho la labor.

Se verifica el siguiente resultado:

**Lema 6.5**

Si $A \in \mathcal{M}_n(\mathbb{R})$ y $\lambda \in \mathbb{C}$ es un autovalor entonces $\overline{\lambda}$ (su complejo conjugado) también es autovalor y además

$$\overline{A - \lambda I} = A - \overline{\lambda} I$$

La demostración es trivial y se deja al lector.

Por tanto, como resolver el sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$ es equivalente a resolver $\overline{(A - \lambda I)\mathbf{x}} = \mathbf{0}$ y

$$\overline{(A - \lambda I)\mathbf{x}} = \overline{(A - \lambda I)}\,\overline{\mathbf{x}},$$

entonces, por el Lema anterior, las soluciones de $(A - \overline{\lambda} I)\mathbf{x} = \mathbf{0}$ son las conjugadas de las soluciones de $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

**Ejemplo 6.11 (continuación del ejemplo 6.10)**

En nuestro caso, como $\lambda = -i$ es el conjugado de $\lambda = i$, entonces los vectores de $E_1(-i)$ coinciden con los conjugados de $E_1(i)$. Luego una base de $E_1(-i)$ estará formada por $(-i, 0, 1, 0)$.

Análogamente, puesto que $\overline{(A - \lambda I)^2} = (A - \overline{\lambda} I)^2$ los vectores de $E_2(-i)$ son los conjugados de los vectores de $E_2(i)$. Así pues, una base de $E_2(-i)$ será

$$\{(-i, 0, 1, 0), (-1, 1+i, 0, 2)\}$$

La partición de multiplicidades es claramente la misma para ambos autovalores, lo que significa que la aportación del autovalor $\lambda = -i$ a la forma de

Jordan es:

$$\begin{pmatrix} -i & 1 \\ 0 & -i \end{pmatrix}$$

Y las mismas operaciones anteriores son válidas para los vectores asociados, en la versión conjugada, es decir

| | | |
|---|---|---|
| $p_2 = 1$ | $E_2(-i)$ | $\mathbf{u}_1^2 = (-1, 1+i, 0, 2)$ |
| $p_1 = 1$ | $E_1(-i)$ | $\mathbf{u}_1^1 = (1, 0, i, 0)$ |

Finalmente, la forma de Jordan y la matriz de paso se escriben

$$J = \left(\begin{array}{cc|cc} i & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ \hline 0 & 0 & -i & 1 \\ 0 & 0 & 0 & -i \end{array}\right) \qquad P = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1-i & 0 & 1+i \\ -i & 0 & i & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix}$$

Como conclusión podemos afirmar que en el caso de autovalores complejos de una matriz real, puesto que éstos vienen por pares conjugados, sólo es necesario hacer el estudio sobre uno de ellos, obteniéndose los conjugados de los resultados obtenidos para el otro.

### 6 3 1 Forma de Jordan real

Como hemos visto en la sección anterior, una matriz real $A \in \mathcal{M}_n(\mathbb{R})$ puede tener autovalores complejos que hagan que su forma de Jordan sea una matriz compleja. En esta sección veremos como asociar a una matriz real una pseudo-forma de Jordan también real, que por abuso de lenguaje llamaremos forma de Jordan real.

Para construir esta nueva matriz debemos recordar que el Lema 6.5 afirma que los autovalores de una matriz real vienen por pares conjugados y como consecuencia, las particiones de multiplicidad son las mismas.

Entonces, para calcular la forma de Jordan real tendremos en cuenta:

(I) Si $\lambda \in \mathbb{R}$ es un autovalor de $A \in \mathcal{M}_n(\mathbb{R})$, las cajas correspondientes y los vectores asociados permanecen igual.

(II) Si $\lambda = \alpha + \beta i \in \mathbb{C} \setminus \mathbb{R}$ es un autovalor de $A \in \mathcal{M}_n(\mathbb{R})$ (y por tanto $\overline{\lambda} = \alpha - \beta i$ también es autovalor) con $p_t$ cajas de tamaño $t$, $p_{t-1} - p_t$ cajas de tamaño

$t-1$, etc., entonces al par de autovalores $\lambda$ y $\overline{\lambda}$ le asignamos

$$\begin{cases} p_t & \text{cajas de tamaño} & 2t, \\ p_{t-1}-p_t & \text{''} & 2(t-1), \\ p_{t-2}-p_{t-1} & \text{''} & 2(t-2), \\ \vdots & \vdots & \vdots \\ p_1-p_2 & \text{''} & 2. \end{cases}$$

de manera que cada caja tiene la forma

$$\begin{pmatrix} B & I_2 & & \\ & B & I_2 & \\ & & \ddots & \\ & & & B \end{pmatrix} \tag{6.11}$$

donde $B = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ e $I_2$ es la matriz identidad de orden dos.

(III) Si $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ es el conjunto de los vectores que aporta el autovalor $\lambda$ a la forma de Jordan,[8] en el orden adecuado, entonces el par de autovalores $\lambda$, $\overline{\lambda}$ aporta los vectores

$$\{\mathrm{Re}(\mathbf{v}_1), \mathrm{Im}(\mathbf{v}_1), \dots, \mathrm{Re}(\mathbf{v}_r), \mathrm{Im}(\mathbf{v}_r)\} \tag{6.12}$$

donde recordemos que, si escribimos $\mathbf{v}_i = \mathbf{x}_i + i\mathbf{y}_i$ entonces $\mathrm{Re}(\mathbf{v}_i) = \mathbf{x}_i$ y $\mathrm{Im}(\mathbf{v}_i) = \mathbf{y}_i$.

**Proposición 6.6**

Con las notaciones anteriores, los vectores de (6.12) pertenecen al espacio $E_t(\lambda) \oplus E_t(\overline{\lambda})$, son linealmente independientes y dan lugar a la caja (6.11).

*Demostración:*
Es evidente que los vectores de (6.12) están en $E_t(\lambda) \oplus E_t(\overline{\lambda})$ pues son combinación lineal de los vectores de estos subespacios:

$$\mathrm{Re}(\mathbf{v}_j) = \frac{1}{2}\mathbf{v}_j + \frac{1}{2}\overline{\mathbf{v}}_j, \quad \mathrm{Im}(\mathbf{v}_j) = -\frac{i}{2}\mathbf{v}_j + \frac{i}{2}\overline{\mathbf{v}}_j \tag{6.13}$$

[8] Lo que implica que $\overline{\mathbf{v}}_1, \dots, \overline{\mathbf{v}}_r$ son los vectores que aporta $\overline{\lambda}$.

Además, son linealmente independientes pues si consideramos la combinación lineal

$$a_1 \operatorname{Re}(\mathbf{v}_1) + b_1 \operatorname{Im}(\mathbf{v}_1) + \cdots + a_r \operatorname{Re}(\mathbf{v}_r) + b_r \operatorname{Im}(\mathbf{v}_r) = \mathbf{0}$$

usando (6.13),

$$\left(\frac{a_1}{2} - \frac{ib_1}{2}\right)\mathbf{v}_1 + \left(\frac{a_1}{2} + \frac{ib_1}{2}\right)\overline{\mathbf{v}}_1 + \cdots + \left(\frac{a_r}{2} - \frac{ib_r}{2}\right)\mathbf{v}_r + \left(\frac{a_r}{2} + \frac{ib_r}{2}\right)\overline{\mathbf{v}}_r = \mathbf{0}$$

y como los vectores $\mathbf{v}_j$ y $\overline{\mathbf{v}}_j$ son linealmente independientes, los coeficientes son nulos, esto es,

$$\frac{a_1}{2} - \frac{ib_1}{2} = \frac{a_1}{2} + \frac{ib_1}{2} = \cdots \frac{a_r}{2} - \frac{ib_r}{2} = \frac{a_r}{2} + \frac{ib_r}{2} = 0$$

de donde $a_j = b_j = 0$, para $1 \leq j \leq r$.

Finalmente, dado que $A\mathbf{v}_j = (\alpha + i\beta)\mathbf{v}_j$, cuando $\mathbf{v}_j$ es un autovector, o bien, $A\mathbf{v}_j = (\alpha + i\beta)\mathbf{v}_j + \mathbf{v}_{j-1}$, separando partes reales e imaginarias se tiene:

$$A \operatorname{Re}(\mathbf{v}_j) = \alpha \operatorname{Re}(\mathbf{v}_j) - \beta \operatorname{Im}(\mathbf{v}_j), \quad A \operatorname{Im}(\mathbf{v}_j) = \beta \operatorname{Re}(\mathbf{v}_j) + \alpha \operatorname{Im}(\mathbf{v}_j)$$

en el primer caso; o bien

$$A \operatorname{Re}(\mathbf{v}_j) = \alpha \operatorname{Re}(\mathbf{v}_j) - \beta \operatorname{Im}(\mathbf{v}_j) + \operatorname{Re}(\mathbf{v}_{j-1})$$
$$A \operatorname{Im}(\mathbf{v}_j) = \beta \operatorname{Re}(\mathbf{v}_j) + \alpha \operatorname{Im}(\mathbf{v}_j) + \operatorname{Im}(\mathbf{v}_{j-1})$$

en el segundo, lo que conduce a cajas de la forma (6.11).

■

**Ejemplo 6.12**

En el ejemplo 6.11 obtuvimos como forma de Jordan (compleja) y matriz de paso las matrices:

$$J = \begin{pmatrix} i & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 1 \\ 0 & 0 & 0 & -i \end{pmatrix} \qquad P = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1-i & 0 & 1+i \\ -i & 0 & i & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix}$$

Por tanto, teniendo en cuenta que $\lambda = i \Rightarrow \alpha = 0$, $\beta = 1$, la forma de Jordan real y la correspondiente matriz de paso se escriben

$$J_{\mathbb{R}} = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{array}\right) \qquad P_{\mathbb{R}} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

es decir, como al autovalor complejo (y a su conjugado) le corresponde una caja de tamaño dos, ahora al par de autovalores les corresponden un bloque de tamaño cuatro con la estructura (6.11).

### 6 3 2 Cálculo de potencias de matrices

Una de las ventajas que supone trabajar con la forma canónica de Jordan es que permite simplificar cierto tipo de cálculo, como puede ser el cálculo de la potencia $k$-ésima de una matriz. Más concretamente, si queremos calcular $A^k$ y conocemos la forma de Jordan y la matriz de paso tenemos que $A = PJP^{-1}$. Entonces

$$A^k = \left(PJP^{-1}\right)^k = \left((PJP^{-1})(PJP^{-1}) \overset{k)}{\cdots} (PJP^{-1})\right) = PJ^kP^{-1}$$

de manera que para conocer la potencia $k$-ésima de $A$ sólo es necesario conocer la correspondiente potencia de $J$ y multiplicar convenientemente por la matriz de paso. La ventaja es que $J$ es una matriz mucho más sencilla que $A$.

Para calcular $J$ se pueden presentar los siguientes casos:

- $J$ diagonal,

$$J = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \Longrightarrow J^k = \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix}$$

- $J$ no diagonal,

$$J = \begin{pmatrix} \lambda_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_n \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & 0 & & \\ & \ddots & \ddots & \\ & & \ddots & 0 \\ & & & \lambda_n \end{pmatrix} + \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

de modo que,

$$J^k = \left[\begin{pmatrix} \lambda_1 & 0 & & \\ & \ddots & \ddots & \\ & & \ddots & 0 \\ & & & \lambda_n \end{pmatrix} + \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}\right]^k = (D+N)^k$$

con

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \quad \text{y} \quad N = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

La ventaja de esta descomposición es que $D$ es una matriz diagonal, de modo que sus potencias son fáciles de calcular, mientras que $N$ es una matriz nilpotente de orden $n$, esto es, $N^n$ es la matriz nula. Así, usando el binomio de Newton para matrices (pues $D$ y $N$ conmutan),

$$J^k = (D+N)^k = \sum_{i=0}^{k} \binom{k}{i} D^{k-i} N^i = \sum_{i=0}^{n-1} \binom{k}{i} D^{k-i} N^i$$

es decir, la suma anterior sólo llega hasta $i = n-1$, pues a partir de $N^n$, el resto de potencias es nulo. Es más, el cálculo de las potencias de $N$ resulta sencillo:

$$\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 1 & & \\ & 0 & 0 & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots & 1 \\ & & & & & 0 \end{pmatrix}$$

esto es, con cada potencia la fila de 1 se desplaza hacia arriba.

**Ejemplo 6.13**

Calculemos $A^k$ donde

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$$

Realizando lo cálculos oportunos se tiene que los autovalores son $\lambda = 3$ (doble), de manera que

$$J = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

y una matriz de paso es

$$P = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

Así pues, $A^k = PJ^kP^{-1}$, y

$$\begin{aligned} J^k &= \left[\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right]^k = \sum_{i=0}^{1} \binom{k}{i} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^i \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}^{k-i} \\ &= \binom{k}{0} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^0 \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}^k + \binom{k}{1} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}^{k-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3^k & 0 \\ 0 & 3^k \end{pmatrix} + k \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3^{k-1} & 0 \\ 0 & 3^{k-1} \end{pmatrix} = \begin{pmatrix} 3^k & k3^{k-1} \\ 0 & 3^k \end{pmatrix} \end{aligned}$$

Nótese que $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^j$ es la matriz nula para $j \geq 2$. Así,

$$A^k = PJ^kP^{-1} = \begin{pmatrix} 3^{k-1}(3-k) & k3^{k-1} \\ -k3^{k-1} & 3^{k-1}(3+k) \end{pmatrix}$$

## 6.4 CÁLCULO CON PYTHON

Como el lector habrá podido comprobar, en este tema se realizan gran cantidad de cálculos que involucran matrices de cierto tamaño, por lo que el uso de Python está más que justificado. Aunque vamos a ver funciones específicas que posee Python para el cálculo de autovectores, autovalores, etc., es posible realizar estos cálculos con lo que ya conocemos de Python.

Dado que queremos realizar cálculos exactos, el módulo SymPy jugará un papel fundamental en esta sección. Comencemos calculando el polinomio carac-

terístico de la matriz

$$A = \begin{pmatrix} -1 & -4 & -3 & -2 & -1 \\ 0 & -13 & -10 & -7 & -3 \\ 0 & -85 & -59 & -40 & -22 \\ 0 & 104 & 74 & 50 & 27 \\ 0 & 90 & 62 & 43 & 22 \end{pmatrix}$$

que resulta:

$$|A - \lambda I| = 3 + 11x + 14x^2 + 6x^3 - x^4 - x^5$$

y cuyas raíces son $\lambda = -1$ de multiplicidad 4 y $\lambda = 3$ de multiplicidad 1:

```
>>> from numpy import matrix
>>> from sympy import Matrix,symbols,roots,eye
>>> a=matrix('-1 -4 -3 -2 -1; 0 -13 -10 -7 -3;
0 -85 -59 -40 -22;0 104 74 50 27;0 90 62 43 22')
>>> A=Matrix(a)
>>> x=symbols('x')
>>> C=A-x*eye(5)
>>> p=C.det()
>>> p
3 + 11*x + 14*x**2 + 6*x**3 - x**4 - x**5

>>> roots(p)
{-1: 4, 3: 1}
```

Como hemos hecho en ocasiones anteriores, hemos construido la matriz usando la función de NumPy, puesto que es más cómoda de manejar. Por su parte, hemos definido un símbolo `x` (en lugar de $\lambda$) con el que construir la expresión simbólica $A - \lambda I$ (nótese el uso de la función `eye` para construir la matriz identidad) y luego hemos calculado el determinante. La función `roots` nos permite obtener las raíces del mismo, junto sus multiplicidades. Nótese que el resultado es un diccionario, cuyas claves son las raíces, y sus valores asociados, las multiplicidades correspondientes.

Podemos calcular ahora los subespacios asociados a cada autovalor:

```
>>> B=A+eye(5)
>>> e1=B.nullspace()
>>> e1
[[1]
[0]
[0]
[0]
[0]]
>>> e2=(B**2).nullspace()
>>> e2
```

```
[[1]
[0]
[0]
[0]
[0], [   0]
[-1/4]
[-7/8]
[ 5/4]
[   1]]
>>> e3=(B**3).nullspace()
>>> e3
[[1]
[0]
[0]
[0]
[0], [    0]
[ -1/5]
[-3/10]
[    1]
[    0], [   0]
[   0]
[-1/2]
[   0]
[   1]]
>>> e4=(B**4).nullspace()
>>> e4
[[1]
[0]
[0]
[0]
[0], [    0]
[-6/11]
[    1]
[    0]
[    0], [    0]
[-4/11]
[    0]
[    1]
[    0], [    0]
[-3/11]
[    0]
[    0]
[    1]]
```

Lo que significa que:

$$E_1(-1) = L((1,0,0,0,0))$$
$$E_2(-1) = L((1,0,0,0,0),(0,-\tfrac{1}{4},-\tfrac{7}{8},\tfrac{5}{4},1))$$
$$E_3(-1) = L((1,0,0,0,0),(0,-\tfrac{1}{5},-\tfrac{3}{10},1,0),(0,0,-\tfrac{1}{2},0,1))$$
$$E_4(-1) = L((1,0,0,0,0),(0,-\tfrac{6}{11},1,0,0),(0,-\tfrac{4}{11},0,1,0),(0,-\tfrac{3}{11},0,0,1)$$

de modo que $\dim(E_1(-1)) = 1$, $\dim(E_2(-1)) = 2$, $\dim(E_3(-1)) = 3$ y $\dim(E_4(-1)) = 4$. A partir de aquí ya tenemos casi toda la información pertinente para construir la forma de Jordan, esto es, $p_1 = 1$, $p_2 = 1$, $p_3 = 1$ y $p_4 = 1$.

Vamos a introducir algunas funciones nuevas para manejar la información obtenida y no tener que reescribirla en cada momento. Por ejemplo, para comprobar qué vector de $E_4(-1)$ es independiente con los de $E_3(-1)$ hemos de construir una matriz, y calcular rangos:

```
>>> from mimodulo import rango
>>> m=e3[0].row_join(e3[1]).row_join(e3[2])
>>> m
[1,     0,    0]
[0,  -1/5,    0]
[0, -3/10, -1/2]
[0,     1,    0]
[0,     0,    1]
>>> rango(m.row_join(e4[3])
4
```

es decir,

$$\operatorname{rango}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{1}{5} & 0 & -\frac{3}{11} \\ 0 & -\frac{3}{10} & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = 4$$

Luego $(0,-\frac{3}{11},0,0,1) \in E_4(-1)\backslash E_3(-1)$.

Obsérvese el uso del atributo `row_join`, que sirve para concatenar una matriz (o vector) con otra. También hemos usado la función `rango` de nuestro propio módulo construida en el tema 5. A partir de aquí, los cálculos son directos:

```
>>> u4=11*e4[3]
>>> u3=B*u4
>>> u2=B*u3
>>> u1=B*u2
>>> P=u1.row_join(u2).row_join(u3).row_join(u4)
```

```
>>> P
[5,   -4,    1,   0]
[0, -10,    3,  -3]
[0, -35,   13,   0]
[0,  50,  -15,   0]
[0,  40,  -17,  11]
```

lo que nos permite construir la tabla

| | | |
|---|---|---|
| $p_4 = 1$ | $E_4(-1)$ | $\mathbf{u}_1^4 = (0, -3, 0, 0, 11)$ |
| $p_3 = 1$ | $E_3(-1)$ | $\mathbf{u}_1^3 = (1, 3, 13, -15, -17)$ |
| $p_2 = 1$ | $E_2(-1)$ | $\mathbf{u}_1^2 = (-4, -10, -35, 50, 40)$ |
| $p_1 = 1$ | $E_1(-1)$ | $\mathbf{u}_1^1 = (5, 0, 0, 0, 0)$ |

Hemos tenido la precaución de multiplicar el primer vector por 11 para evitar en la medida de lo posible las tediosas fracciones. Si añadimos la información proveniente del otro autovalor,

```
>>> B=A-3*eye(5)
>>> u5=B.nullspace()
>>> P=P.row_join(u5[0])
>>> P
[5,   -4,    1,   0,   0]
[0, -10,    3,  -3,   0]
[0, -35,   13,   0,  -1]
[0,  50,  -15,   0,   1]
[0,  40,  -17,  11,   1]
```

nos resulta la matriz de paso:

$$P = \begin{pmatrix} 5 & -4 & 1 & 0 & 0 \\ 0 & -10 & 3 & -3 & 0 \\ 0 & -35 & 13 & 0 & -1 \\ 0 & 50 & -15 & 0 & 1 \\ 0 & 40 & -17 & 11 & 1 \end{pmatrix}$$

Así pues, la forma de Jordan tendrá una caja de tamaño 4 para el autovalor $-1$ y una caja de tamaño 1 para el autovalor 3, algo que podemos comprobar fácilmente

```
>>> J=P*A*P.inv()
>>> J
[-1,   1,   0,   0, 0]
[ 0,  -1,   1,   0, 0]
[ 0,   0,  -1,   1, 0]
[ 0,   0,   0,  -1, 0]
[ 0,   0,   0,   0, 3]
```

Como comentamos al inicio de la sección, tanto NumPy como SymPy poseen funciones destinadas a calcular autovalores y autovectores de forma más rápida. Por otra parte, hay que tener en cuenta que NumPy realizará el cálculo de autovalores y autovectores usando métodos numéricos que dan lugar a las habituales aproximaciones numéricas.

En NumPy podemos calcular autovalores con el módulo `linalg.eigvals`:

```
>>> from numpy import linalg
>>> linalg.eigvals(a)
array([-1.00000000 +0.00000000e+00j,  3.00000000 +0.00000000e+00j,
       -0.99997160 +4.92035100e-05j, -0.99997160 -4.92035100e-05j,
       -1.00005681 +0.00000000e+00j])
```

Nótese que no reconoce la multiplicidad del autovalor $-1$ correctamente. El módulo `linalg.eig(a)` proporciona a la vez, la información de autovalores y los autovectores asociados. Con SymPy, el método es similar: tenemos módulos para vectores y autovectores por separado:

```
>>> A.eigenvals()
{-1: 4, 3: 1}
>>> A.eigenvects()
[(-1, 4, [[1]
[0]
[0]
[0]
[0]]), (3, 1, [[ 0]
[ 0]
[-1]
[ 1]
[ 1]])]
```

aunque como podemos observar, `eigenvects` proporciona también los autovalores, su multiplicidad y los vectores del subespacio propio correspondiente.

En la última versión de SymPy hasta la fecha (`0.7.1`) se acaba de implementar el cálculo de la forma de Jordan, aunque no el de la matriz de paso:

```
>>> A.jordan_form()
(None, [-1,  1,  0,  0, 0]
[ 0, -1,  1,  0, 0]
[ 0,  0, -1,  1, 0]
[ 0,  0,  0, -1, 0]
[ 0,  0,  0,  0, 3])
```

6 5

## APLICACIÓN: OSCILADORES ACOPLADOS

Entre las múltiples aplicaciones de los autovalores y los autovectores hay una muy destacada que tiene que ver con el estudio de vibraciones en sistemas mecánicos. Aquí vamos a mostrar un típico ejemplo no trivial de vibración de un sistema de muelles como el de la figura 6.2, en el que tenemos dos masas iguales conectadas entre sí, y conectadas a su vez a la pared.

Denotemos por $x_1(t)$ y $x_2(t)$ a las funciones que indican el desplazamiento lateral (positivo a la derecha y negativo a la izquierda) de cada una de las masas desde su posición de equilibrio. La fuerza ejercida sobre cada una de las masas (ignorando fuerzas de fricción y fuerzas externas como la gravedad) seguirá la segunda ley de Newton: $F = m \cdot a$, con $m$ la masa y $a$ la aceleración. Por otro lado, la ley de Hooke nos dice que la fuerza que ejerce cada muelle es proporcional a la compresión de éste. Como es bien conocido, la aceleración corresponde a la derivada segunda del desplazamiento, de manera que la fuerza que actúa sobre la primera masa corresponderá a la ejercida por los muelles a los que está conectada. Si desplazamos a la derecha la primera masa, el muelle de la izquierda ejerce una fuerza opuesta al sentido de movimiento igual a $-kx_1$ mientras que el muelle central ejerce una fuerza igual a $k(x_2 - x_1)$, es decir,

$$mx_1''(t) = -kx_1(t) + k(x_2(t) - x_1(t)) = k(-2x_1(t) + x_2(t))$$

donde $k$ es la constante de proporcionalidad del muelle y $m$ la masa. En la segunda masa, un razonamiento similar nos lleva a:

$$mx_2''(t) = -k(x_2(t) - x_1(t)) + k(-x_2(t)) = k(x_1(t) - 2x_2(t))$$

Es decir, tenemos en este caso un sistema de ecuaciones diferenciales de segundo orden (hay derivadas de orden dos), que admite la representación matricial siguiente:

$$\begin{pmatrix} x_1'' \\ x_2'' \end{pmatrix} = \frac{k}{m} \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

donde hemos omitido la dependencia en $t$ para simplificar la notación. En forma vectorial este sistema se escribe

$$\mathbf{x}'' = A\mathbf{x} \quad \text{con } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ y } A = \frac{k}{m} \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

No vamos a entrar aquí en cómo resolver sistemas de ecuaciones diferenciales como éste, aunque hemos de resaltar que la forma de Jordan de la matriz del sistema juega un papel fundamental en la resolución. Lo que haremos será probar con cierto tipo de soluciones expresadas a través de la exponencial compleja: en concreto, buscaremos soluciones del sistema de la forma $x_1(t) = c_1 e^{i\omega t}$ y

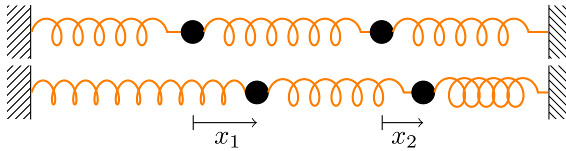



Figura 6.2: Sistema de muelles

$x_2(t) = c_2 e^{i\omega t}$, donde $c_1$, $c_2$ y $\omega$ son constantes a determinar. Derivando y sustituyendo en el sistema:

$$\begin{cases} c_1 i^2\omega^2 e^{i\omega t} = \frac{k}{m}\left(-2c_1 e^{i\omega t} + c_2 e^{i\omega t}\right) \\ c_2 i^2\omega^2 e^{i\omega t} = \frac{k}{m}\left(c_1 e^{i\omega t} - 2c_2 e^{i\omega t}\right) \end{cases} \Rightarrow \begin{cases} -c_1\omega^2 e^{i\omega t} = \frac{k}{m}(-2c_1 + c_2)e^{i\omega t} \\ -c_2\omega^2 e^{i\omega t} = \frac{k}{m}(c_1 - 2c_2)e^{i\omega t} \end{cases}$$

que simplificando y agrupando factores resulta

$$\begin{pmatrix} 2\frac{k}{m} - \omega^2 & -\frac{k}{m} \\ -\frac{k}{m} & 2\frac{k}{m} - \omega^2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Este sistema tendrá soluciones no triviales, esto es, distintas de la solución nula, si

$$\begin{vmatrix} 2\frac{k}{m} - \omega^2 & -\frac{k}{m} \\ -\frac{k}{m} & 2\frac{k}{m} - \omega^2 \end{vmatrix} = 0$$

Ahora bien, la condición anterior se puede escribir como

$$\det\left(\frac{k}{m}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} - \omega^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = 0$$

es decir, $\omega^2$ coincide con los autovalores de la matriz $A$. Y para cada uno de estos autovalores las constantes $c_1$ y $c_2$ son las componentes de los autovectores asociados a cada autovalor. Un simple cálculo nos muestra que los autovalores y autovectores resultan:

$$\omega_1 = \sqrt{\frac{k}{m}} \to (1, -1), \quad \omega_2 = \sqrt{\frac{3k}{m}} \to (1, 1)$$

Cualquier otra solución se va a poder expresar como una combinación lineal de las dos soluciones obtenidas, denominadas *modos normales* de vibración.

Los modos normales de vibración corresponden a soluciones del oscilador en el que ambas masas se mueven con igual frecuencia. La figura 6.3 muestra el movimiento del sistema de ambos modos. Obsérvese que el autovector asociado a la frecuencia $\omega_1$, $(1, -1)$ nos indica que el movimiento de cada una de las

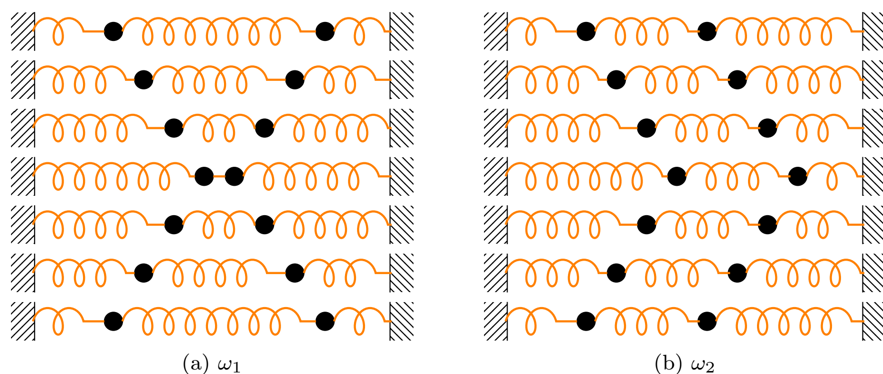



Figura 6.3: Movimiento de los modos normales

masas tiene dirección opuesta (figura 6.3a), mientras que para la frecuencia $\omega_2$, se mueven en la misma dirección (figura 6.3b).

A la vista de los resultados, ¿cuál será el modo de vibración más peligroso para la estructura? Teniendo en cuenta que con la frecuencia $\omega_1$ se producen los mayores estiramientos y compresiones de los muelles, parece claro que este modo de vibrar puede causar más daño en la estructura que cualquier otro. Nótese además que este modo, frecuentemente llamado *modo fundamental*, corresponde a la frecuencia asociada al *menor autovalor*.

En el estudio de estructuras es importante calcular el modo fundamental de éstas, pues si son excitadas con una frecuencia próxima a su modo fundamental los desplazamientos producidos pueden superar el máximo de lo que la propia estructura puede resistir. Tal y como hemos visto, el cálculo del modo fundamental corresponde al menor de los autovalores.

6 6

## EJERCICIOS

### Ejercicios de repaso

**E.1** Hallar el polinomio característico y los autovalores de las siguientes matrices:

(a) $\begin{pmatrix} 4 & 6 \\ -3 & -5 \end{pmatrix}$ (b) $\begin{pmatrix} 5 & -1 \\ 4 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} -1 & -5 & -4 \\ 6 & 10 & 6 \\ -6 & -7 & -3 \end{pmatrix}$ (d) $\begin{pmatrix} 0 & 0 & -1 \\ 1 & -2 & -1 \\ -2 & 3 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 8 & -6 & -2 \\ 18 & -13 & -3 \\ 6 & -3 & -5 \end{pmatrix}$ (f) $\begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix}$

(g) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ (h) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix}$

**E.2** Para cada una de las matrices del ejercicio 1 hallar, si es posible, una base de autovectores.

**E.3** Decir cuáles de las siguientes matrices pueden reducirse a una diagonal y encontrar una matriz de cambio de base (real):

(a) $\begin{pmatrix} -1 & -3 & 1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 2 & 2 & -1 \\ 0 & 2 & 0 \\ 3 & 1 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 4 & -1 & -1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}$

(d) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ (e) $\begin{pmatrix} 6 & 6 & 4 & 4 \\ -4 & 2 & 0 & -4 \\ 0 & -2 & -2 & 2 \\ -4 & 0 & 0 & -2 \end{pmatrix}$

**E.4** Encontrar la forma canónica de Jordan y la matriz de cambio de base correspondiente de las matrices

(a) $\begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} -14 & 1 & 12 \\ -13 & 0 & 12 \\ -17 & 1 & 15 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & -1 & -2 \\ 1 & 3 & 1 \\ 1 & 0 & 3 \end{pmatrix}$ (d) $\begin{pmatrix} 3 & 2 & -3 \\ 4 & 10 & -12 \\ 3 & 6 & -7 \end{pmatrix}$

**E.5** En los siguientes casos hallar los autovalores (reales o complejos), los autovectores de $\mathbb{C}^n$ y la matriz de paso en su forma real y/o compleja de:

(a) $\begin{pmatrix} 1 & 1+i \\ -i & 1-i \end{pmatrix}$ (b) $\begin{pmatrix} 1 & -1 \\ 9 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

**E.6** Calcular $A^{50}$ donde

$$A = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & -2 \\ 2 & 2 & -3 \end{pmatrix}$$

## Problemas

**E.7** Determinar los coeficientes $a$, $b$, $c$, $d$, $e$ y $f$ de la matriz

$$A = \begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ d & e & f \end{pmatrix}$$

sabiendo que $(1,1,1)$, $(1,0,-1)$ y $(1,-1,0)$ son autovectores. ¿Cuáles son los autovalores de la matriz?

**E.8** Determinar para qué valores $a, b \in \mathbb{R}$, la matriz

$$A = \begin{pmatrix} a & b & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathcal{M}_3(\mathbb{R})$$

es diagonalizable.

**E.9** Dada la matriz

$$A = \begin{pmatrix} 2 & -2 & 6 \\ 0 & a & 4-a \\ 0 & a & -a \end{pmatrix}$$

probar que ésta es diagonalizable para todo $a \neq 0$.

**E.10** Hallar las formas canónicas y las correspondientes matrices de paso de las aplicaciones $f$ y $g$ cuyas matrices respecto de la base canónica son, respectivamente,

$$\text{(a)} \begin{pmatrix} -1 & -4 & -3 & -2 & -1 \\ 0 & -13 & -10 & -7 & -3 \\ 0 & -85 & -59 & -40 & -22 \\ 0 & 104 & 74 & 50 & 27 \\ 0 & 90 & 62 & 43 & 22 \end{pmatrix} \qquad \text{(b)} \begin{pmatrix} 5 & -4 & -3 & -2 & -1 \\ 0 & -12 & -10 & -8 & -3 \\ 0 & -13 & -11 & -13 & -1 \\ 0 & 37 & 32 & 30 & 6 \\ 0 & 28 & 20 & 18 & 7 \end{pmatrix}$$

Indicación: $P_f(\lambda) = (\lambda+1)^4(3-\lambda)$ y $P_g(\lambda) = (5-\lambda)^3(\lambda-2)^2$.

**E.11** Hallar la potencia $n$-ésima de la matriz

$$A = \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix}$$

**E.12** Calcular, si es posible, el $\lim\limits_{n\to\infty} A^n$, donde $A$ es la matriz

(a) $\begin{pmatrix} 0.9 & 0.15 \\ 0.1 & 0.85 \end{pmatrix}$ (b) $\begin{pmatrix} 0.5 & 0 \\ 0.5 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} \frac{5}{2} & 3 & 0 \\ -\frac{3}{2} & -2 & 0 \\ -6 & -6 & -\frac{1}{2} \end{pmatrix}$

### Ejercicios teóricos

**E.13** Probar que los autovalores de una matriz real y simétrica de orden dos son siempre reales.

**E.14** Probar que si $A$ es no singular, entonces los autovalores de $A^{-1}$ son los inversos de los autovalores de $A$.

***E.15** Encontrar los autovalores y autovectores de la aplicación derivada

$$\begin{array}{rccc} D: & \mathbb{P}^3_{\mathbb{R}}[x] & \to & \mathbb{P}^3_{\mathbb{R}}[x] \\ & p(x) & \to & p'(x) \end{array}$$

**E.16** Probar que si $\lambda$ es un autovalor de una matriz $A$ asociado a un autovector $\mathbf{x}$ y $\mu$ es un autovalor de otra matriz $B$ asociado al mismo autovector $\mathbf{x}$, entonces $\lambda + \mu$ es un autovalor de $A + B$ asociado a $\mathbf{x}$.

**E.17** Sean $A, B \in \mathcal{M}_n(\mathbb{K})$ dos matrices invertibles, $\mathbf{v} \in \mathbb{K}^n$ un vector no nulo y $\lambda \in \mathbb{K}$ un escalar no nulo tales que se verifica $A\mathbf{v} = \lambda B\mathbf{v}$. ¿Cuáles de las siguientes afirmaciones son ciertas?

(a) $\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $A^{-1}B$.

(b) $\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $BA^{-1}$.

(c) $\mathbf{v}$ es un autovector asociado al autovalor $\lambda^{-1}$ de la matriz $BA^{-1}$.

(d) $\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $AB^{-1}$.

(e) $\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $B^{-1}A$.

***E.18** Sean $A, B \in \mathcal{M}_n(\mathbb{K})$. ¿Cuáles de las siguientes afirmaciones son ciertas?

(a) $A$ es diagonalizable si y solo si $A - I$ es diagonalizable.

(b) 1 es autovalor de $A$ si y sólo si $A - I$ no es invertible.

(c) Si $\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $BA$ y $B$ es invertible, entonces $B^{-1}\mathbf{v}$ es un autovector asociado al autovalor $\lambda$ de la matriz $AB$.

(d) 0 es autovalor de $AB$ si y sólo si 0 es autovalor de $BA$.

*E.19 Dada una suma directa $V = W \oplus Z$ podemos definir $E \in \mathcal{L}(V)$ como $E(\mathbf{v}) = \mathbf{z}$ si $\mathbf{v} = \mathbf{z} + \mathbf{w}$ con $\mathbf{z} \in Z$ y $\mathbf{w} \in W$.

(a) Probar que $E^2 = E$, $W = \ker(E)$ y $Z = \operatorname{Im}(E)$.

(b) Encontrar los autovalores de la aplicación $E$.

*E.20 Supongamos que $A$ es una matriz equivalente a otra matriz $D$, diagonal donde los elementos de la diagonal son $\lambda_1, \lambda_2, ..., \lambda_n$ tal que $|\lambda_i| < 1$ para $i = 1, ..., n$. Sea $p_k(A) = I + A + A^2 + ... + A^k$.

(a) Probar que si $A = PDP^{-1}$ entonces

$$p_k(A) = P(I + D + D^2 + \cdots + D^k)P^{-1}$$

(b) Demostrar que $\lim\limits_{k\to\infty} p_k(A)$ existe.

(c) Si denotamos por $B$ a la matriz $\sum_{k=1}^{\infty} A^k = I + A + A^2 + \dots$, probar que $B = (I - A)^{-1}$.

Indicación: considérese $\lim\limits_{k\to\infty} (I - A)p_k(A)$.

*E.21 Supongamos que la matriz $A$ es equivalente a la matriz diagonal $D$ cuyos elementos de la diagonal son $\lambda_1, \lambda_2, ..., \lambda_n$. Sea

$$p_k(A) = I + \frac{A}{1!} + \frac{A^2}{2!} + ... + \frac{A^k}{k!}.$$

Demostrar que si $A = PDP^{-1}$ entonces

$$\lim_{k\to\infty} p_k(A) = P \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix} P^{-1}$$

Indicación: $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$.

## Ejercicios adicionales

**E.22** Usar Python para encontrar la forma de Jordan de la matriz

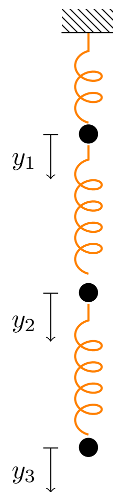
$$A = \begin{pmatrix} -4 & -5 & -3 & 1 & -2 & 0 & 1 & -2 \\ 4 & 7 & 3 & -1 & 3 & 9 & -1 & 2 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 2 & -4 & 2 & 0 & -3 & 1 \\ -8 & -14 & -5 & 1 & -6 & 0 & 1 & 4 \\ 4 & 7 & 4 & -3 & 3 & -1 & -3 & 4 \\ 2 & -2 & -2 & 5 & -3 & 0 & 4 & -1 \\ 6 & 7 & 3 & 0 & 2 & 0 & 0 & 3 \end{pmatrix}$$

**E.23** Considérese un sistema de muelles como el de la figura adjunta compuesto por tres masas iguales conectadas entre sí por tres muelles de idéntica longitud y la misma constante de proporcionalidad $k$.

Denotemos por $y_1$, $y_2$ e $y_3$ al desplazamiento vertical (hacia abajo) de cada una de las masas. Probar que el sistema de ecuaciones diferenciales que modela el movimiento del sistema viene dado por

$$\begin{pmatrix} y_1'' \\ y_2'' \\ y_3'' \end{pmatrix} = \frac{k}{m} \begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

y usar Python para encontrar los modos fundamentales de la estructura.



# 7 Ecuaciones lineales en diferencias

7 1

## INTRODUCCIÓN

Numerosos procesos temporales, esto es, que evolucionan con el tiempo, pueden ser modelados a través de sistemas discretos, en los cuales se considera que el paso de un estado del sistema al siguiente está separado por un intervalo temporal determinado.

Por ejemplo, imaginemos que pretendemos averiguar la evolución de un fondo de inversión cuyos intereses se abonan al final de un período estipulado (usualmente, cada mes). De este modo, el capital existente no se incrementa de forma continua, sino que lo hace mes a mes. Los intereses abonados al final del primer mes incrementan el capital, de manera que el cálculo de los mismos al mes siguiente difiere del cálculo en el mes anterior.

De este modo, podemos establecer una sucesión de valores que proporcione el capital obtenido en cada mes; esto es, $x_0$, $x_1$, $x_2$, ..., siendo $x_0$ nuestro capital inicial.

El cálculo del capital en el mes $k$-ésimo vendrá dado por el capital en el mes $k-1$ junto con los intereses que ha generado. Si $i$ denota el interés mensual en %, entonces

$$x_k = x_{k-1} + \frac{i}{100} x_{k-1} = \left(1 + \frac{i}{100}\right) x_{k-1}, \quad k \geq 1$$

Esta ecuación representa un fenómeno dinámico discreto, y se denomina comúnmente *ecuación en diferencias*. La ecuación proporciona los valores en la etapa $k$ a partir de los valores en la etapa $k-1$. El interés en estas ecuaciones está en averiguar el estado del sistema en una etapa determinada, o su evolución en el estado límite, es decir, cuando $k \to \infty$, sin necesidad de ir recorriendo todos los estados uno a uno.

En este ejemplo concreto, es fácil obtener el valor en la etapa $k$ pues,

$$x_k = \left(1 + \frac{i}{100}\right) x_{k-1} = \left(1 + \frac{i}{100}\right)^2 x_{k-2} = \cdots = \left(1 + \frac{i}{100}\right)^k x_0$$

de manera que es posible obtener dicho valor directamente a partir del dato $x_0$. Es decir, hemos resuelto la ecuación en diferencias.

Este tipo de ecuaciones pueden complicarse, por ejemplo, variando el tipo de interés de un mes a otro. En tal caso, la ecuación resultaría

$$x_k = \left(1 + \frac{i(k)}{100}\right) x_{k-1},$$

con $i(k)$ el interés en el mes $k$-ésimo, lo que obviamente dificulta el estudio.

De forma similar, pueden encontrarse fenómenos modelados por ecuaciones en diferencias de orden superior, esto es, que involucran no sólo lo que ocurre en un estado inmediatamente anterior, sino lo ocurrido en estados anteriores, como por ejemplo, la sucesión de Fibonacci[1]

$$x_{k+2} = x_{k+1} + x_k, \quad k \geq 0$$

En este caso estamos ante una ecuación en diferencias de orden dos (pues el cálculo de un estado involucra dos estados anteriores).

Incluso es posible tener sistemas de ecuaciones en diferencias, más conocidos como *sistemas dinámicos discretos* que equivalen a sucesiones multidimensionales, es decir, una sucesión de valores vectoriales, cuya evolución, al igual que las ecuaciones en diferencias, viene marcada por el comportamiento del sistema en el estado anterior (en los sistemas de primer orden), o en estados anteriores (si el sistema es de orden superior). Al igual que antes, el interés en este tipo de sistemas es estudiar su dinámica, es decir averiguar hacia dónde tiende el sistema a partir de un dato inicial.

**Definición 7.1**

Una *ecuación en diferencias de orden n* es una expresión del tipo

$$f(k, x_k, x_{k+1}, \dots, x_{k+n}) = 0$$

Si $f$ es lineal en $x_k, \dots, x_{k+n}$ se dirá ecuación en diferencias *lineal*, que de forma genérica se escribe

$$a_n(k)x_{k+n} + \cdots + a_1(k)x_{k+1} + a_0(k)x_k = g(k)$$

donde $a_0, \dots, a_n$ y $g$ son funciones. En el caso en el que $a_0, \dots, a_n$ sean constantes se dirá *ecuación en diferencias lineal con coeficientes constantes*. Si además $g = 0$ se dice ecuación *homogénea* y si $g \neq 0$ *no homogénea*.

Si $f$ es una función con valores vectoriales y $x_k$ es también un vector, estaremos hablando de *sistemas de ecuaciones en diferencias*.

---

[1]La sucesión toma su nombre del italiano Leonardo da Pisa, más conocido como Fibonacci, aunque aparece con anterioridad en trabajos de matemáticos indios. Tiene numerosas aplicaciones en ciencias de la computación, biología o teoría de juegos.

**Proposición 7.1**

Si $y_k^{(1)}, \ldots, y_k^{(m)}$ son soluciones de la ecuación lineal homogénea, entonces cualquier combinación lineal suya también es solución.

*Demostración:*
En efecto, si $y_k^{(1)}, \ldots, y_k^{(m)}$ resuelven la ecuación

$$a_n(k)x_{k+n} + \cdots + a_1(k)x_{k+1} + a_0(k)x_k = 0$$

se tendrá:

$$\begin{aligned} a_n(k)y_{k+n}^{(1)} + \cdots + a_1(k)y_{k+1}^{(1)} + a_0(k)y_k^{(1)} &= 0 \\ a_n(k)y_{k+n}^{(2)} + \cdots + a_1(k)y_{k+1}^{(2)} + a_0(k)y_k^{(2)} &= 0 \\ &\vdots \\ a_n(k)y_{k+n}^{(m)} + \cdots + a_1(k)y_{k+1}^{(m)} + a_0(k)y_k^{(m)} &= 0 \end{aligned}$$

Si ahora consideramos una combinación lineal de las ecuaciones anteriores:

$$\begin{aligned} 0 &= c_1\left(a_n(k)y_{k+n}^{(1)} + \cdots + a_1(k)y_{k+1}^{(1)} + a_0(k)y_k^{(1)}\right) + \cdots \\ &\qquad + c_m\left(a_n(k)y_{k+n}^{(m)} + \cdots + a_1(k)y_{k+1}^{(m)} + a_0(k)y_k^{(m)}\right) \\ &= a_n(k)\left(c_1y_{k+n}^{(1)} + \cdots + c_my_{k+n}^{(m)}\right) + \cdots + a_1(k)\left(c_1y_{k+1}^{(1)} + \cdots + c_my_{k+1}^{(m)}\right) \\ &\qquad + a_0(k)\left(c_1y_k^{(1)} + \cdots + c_my_k^{(m)}\right) \end{aligned}$$

es decir, $c_1y_k^{(1)} + \cdots + c_my_k^{(m)}$ es solución de la ecuación en diferencias.

■

**Definición 7.2**

Un conjunto de $n$ soluciones linealmente independientes de una ecuación lineal homogénea de orden $n$ se dice *sistema fundamental* de soluciones.

**Definición 7.3**

Llamaremos *solución particular* a cualquier solución de la ecuación no homogénea.

**Proposición 7.2**

Si $y^{(h)}$ es solución de la ecuación lineal homogénea e $y^{(p)}$ es una solución particular de la ecuación no homogénea, la suma de ambas también es solución de la ecuación no homogénea.

La demostración es elemental y se deja al lector.

## 7 2 ECUACIONES Y SISTEMAS LINEALES DE PRIMER ORDEN

Consideremos la ecuación de primer orden homogénea con coeficientes constantes:

$$a_1 x_{k+1} + a_0 x_k = 0$$

Dividiendo por $a_1$ (si $a_1 = 0$ la solución es trivialmente cero), se tiene que $x_{k+1} = a x_k$, con $a = -\frac{a_0}{a_1}$. Realizando reiteradas sustituciones hacia atrás,

$$x_k = a^k x_0$$

expresión que nos proporciona la solución de la ecuación lineal de coeficientes constantes de primer orden.

Para tratar el caso de sistemas lineales de ecuaciones en diferencias con coeficientes constantes vamos a ver unos cuantos ejemplos en dimensión dos.

**Ejemplo 7.1**

(I) Estudiemos el siguiente sistema de primer orden homogéneo con coeficientes constantes:

$$\begin{cases} x_{k+1} &= x_k - y_k \\ y_{k+1} &= 2x_k + 4y_k \end{cases}$$

El sistema se puede representar en forma matricial,

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

o abreviadamente $\mathbf{x}_{k+1} = A\mathbf{x}_k$, con

$$\mathbf{x}_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}$$

Al igual que en el caso anterior, si realizamos sustituciones regresivas:

$$\mathbf{x}_k = A\mathbf{x}_{k-1} = AA\mathbf{x}_{k-2} = A^3\mathbf{x}_{k-3} = \cdots = A^k\mathbf{x}_0$$

Nótese que para resolver estos sistemas tendríamos que calcular la potencias $k$-ésimas de la matriz $A$, tal y como vimos en la sección 6.3.2. Sin embargo, vamos a aprovechar la información proveniente de la forma de Jordan para obtener la solución del sistema dinámico de modo que no sea necesario calcular inversas y multiplicaciones de matrices.

Calculamos para ello la forma de Jordan de $A$: sus autovalores son $\lambda = 2$ y $\lambda = 3$, y por tanto se trata de una matriz diagonalizable. Si consideramos la base formada por los autovectores, y denotamos por $P$ a la matriz de cambio de base de la canónica a la base formada por los autovectores sabemos que $A = P^{-1}JP$. Si hacemos el cambio de base $\mathbf{y}_k = P\mathbf{x}_k$ entonces

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \Rightarrow P\mathbf{x}_{k+1} = PA\mathbf{x}_k \Rightarrow P\mathbf{x}_{k+1} = JP\mathbf{x}_k$$

y por tanto,

$$\mathbf{y}_{k+1} = J\mathbf{y}_k \Rightarrow \mathbf{y}_k = J^k\mathbf{y}_0$$

Es decir, lo que sucede es que si elegimos la base adecuada (en este caso, la base formada por los autovectores de $A$), el sistema dinámico se puede reescribir en términos de la matriz de Jordan. Puesto que

$$J = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \Rightarrow J^k = \begin{pmatrix} 2^k & 0 \\ 0 & 3^k \end{pmatrix}$$

obtenemos que $\mathbf{y}_k = \begin{pmatrix} c_1 2^k \\ c_2 3^k \end{pmatrix}$, con $c_1$ y $c_2$ las componentes de $\mathbf{y}_0$. Deshaciendo el cambio anterior, $\mathbf{x}_k = P^{-1}\mathbf{y}_k$, se tendrá que

$$\mathbf{x}_k = P^{-1}\begin{pmatrix} c_1 2^k \\ c_2 3^k \end{pmatrix}$$

Observando que $P^{-1}$ es la matriz de cambio de la base de autovectores a la base canónica, bastará calcular los autovectores, que en este caso son $(1, -1)$ y $(1, -2)$ para obtener:

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_1 \\ -c_1 \end{pmatrix} 2^k + \begin{pmatrix} c_2 \\ -2c_2 \end{pmatrix} 3^k$$

como solución general. Las constantes $c_1$ y $c_2$ se determinan a partir de los datos iniciales.

Si miramos el aspecto de la solución general encontrada, podemos evitar de cierta forma la multiplicación de matrices y el cálculo de autovectores razonando del siguiente modo: dado que los autovalores son 2 y 3, la solución general que buscamos tendrá el siguiente aspecto:

$$\mathbf{x}_k = \mathbf{C}_1 2^k + \mathbf{C}_2 3^k$$

donde $\mathbf{C}_1$ y $\mathbf{C}_2$ son vectores constantes a determinar. Dado que se trata de un sistema bidimensional de primer orden, la solución general sólo depende de dos parámetros (y no cuatro, como señalan los vectores $\mathbf{C}_1$ y $\mathbf{C}_2$). Si denotamos por $\mathbf{C}_i = \begin{pmatrix} c_{i1} \\ c_{i2} \end{pmatrix}$ e imponemos cualquiera de las ecuaciones del sistema se obtiene que

$$x_{k+1} = x_k - y_k \Rightarrow c_{11}2^{k+1} + c_{21}3^{k+1} = c_{11}2^k + c_{21}3^k - c_{12}2^k - c_{22}3^k$$

de donde

$$\left\{ \begin{array}{l} 2c_{11} = c_{11} - c_{12} \\ 3c_{21} = c_{21} - c_{22} \end{array} \right. \Rightarrow \left\{ \begin{array}{l} c_{11} = -c_{12} \\ 2c_{21} = -c_{22} \end{array} \right.$$

Es decir, la solución general viene dada por

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_{11} \\ -c_{11} \end{pmatrix} 2^k + \begin{pmatrix} c_{21} \\ -2c_{21} \end{pmatrix} 3^k$$

que coincide con la obtenida antes.

(II) Veamos ahora el sistema

$$\left\{ \begin{array}{rcl} x_{k+1} & = & 2x_k + y_k \\ y_{k+1} & = & -x_k + 4y_k \end{array} \right.$$

que en forma matricial es,

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

El cálculo de la forma de Jordan de $A$ se realizó en el ejemplo 6.12, obteniéndose que

$$J = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \Rightarrow J^k = \begin{pmatrix} 3^k & k3^{k-1} \\ 0 & 3^k \end{pmatrix}$$

Procediendo como en el ejemplo anterior, usando el cambio $\mathbf{y}_k = P\mathbf{x}_k$, donde $P$ será la inversa de la matriz de paso en el cálculo de la forma de Jordan, entonces

$$\mathbf{y}_k = J^k\mathbf{y}_0 \Rightarrow \mathbf{y}_k = \begin{pmatrix} c_1 3^k + c_2 k 3^{k-1} \\ c_2 3^k \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} 3^k + \begin{pmatrix} c_2 \\ 0 \end{pmatrix} k3^{k-1}$$

Para deshacer el cambio, vamos a razonar como en el ejemplo anterior; puesto que $\mathbf{x}_k = P^{-1}\mathbf{y}_k$, en definitiva tendremos que

$$\mathbf{x}_k = \mathbf{C}_1 3^k + \mathbf{C}_2 k 3^k$$

con $\mathbf{C}_1$ y $\mathbf{C}_2$ vectores constantes a determinar (que dependerán solo de dos parámetros, y no de cuatro). Usando que $x_{k+1} = 2x_k + y_k$ y poniendo

$$\mathbf{C}_i = \begin{pmatrix} c_{i1} \\ c_{i2} \end{pmatrix}$$

$$(3c_{11} + 3c_{21})3^k + 3c_{21}k3^k = (2c_{11} + c_{12})3^k + (2c_{21} + c_{22})k3^k$$

se tiene que $c_{12} = c_{11} + 3c_{21}$ y $c_{21} = c_{22}$, por tanto, la solución general queda como

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{11} + 3c_{21} \end{pmatrix} 3^k + \begin{pmatrix} c_{21} \\ c_{21} \end{pmatrix} k3^k$$

(III) Veamos finalmente un ejemplo que involucra autovalores complejos. Dado el sistema de ecuaciones en diferencias

$$\begin{cases} x_{k+1} = x_k + y_k \\ y_{k+1} = -x_k + y_k \end{cases}$$

cuya matriz asociada es $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, su forma de Jordan (compleja) es

$$J = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}$$

En este caso es más conveniente acudir a la forma polar de los números complejos para calcular las potencias de $J$:

$$J^k = \begin{pmatrix} \left(\sqrt{2}\right)^k e^{i\frac{k\pi}{4}} & 0 \\ 0 & \left(\sqrt{2}\right)^k e^{-i\frac{k\pi}{4}} \end{pmatrix}$$

Como $\mathbf{y}_k = J^k \mathbf{y}_0$, obtendremos

$$\mathbf{y}_k = \begin{pmatrix} c_1 \left(\sqrt{2}\right)^k e^{i\frac{k\pi}{4}} \\ c_2 \left(\sqrt{2}\right)^k e^{-i\frac{k\pi}{4}} \end{pmatrix} = \begin{pmatrix} c_1 \left(\sqrt{2}\right)^k \left(\cos\left(\frac{k\pi}{4}\right) + i \operatorname{sen}\left(\frac{k\pi}{4}\right)\right) \\ c_2 \left(\sqrt{2}\right)^k \left(\cos\left(-\frac{k\pi}{4}\right) + i \operatorname{sen}\left(-\frac{k\pi}{4}\right)\right) \end{pmatrix}$$

$$= \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \left(\sqrt{2}\right)^k \cos\left(\frac{k\pi}{4}\right) + \begin{pmatrix} ic_1 \\ -ic_2 \end{pmatrix} \left(\sqrt{2}\right)^k \operatorname{sen}\left(\frac{k\pi}{4}\right)$$

Al igual que en los ejemplos anteriores, multiplicando por $P^{-1}$ llegamos a una solución general de la forma

$$\mathbf{x}_k = \mathbf{C_1} \left(\sqrt{2}\right)^k \cos\left(\frac{k\pi}{4}\right) + \mathbf{C_2} \left(\sqrt{2}\right)^k \operatorname{sen}\left(\frac{k\pi}{4}\right)$$

en la que, para determinar los vectores

$$\mathbf{C}_i = \begin{pmatrix} c_{i1} \\ c_{i2} \end{pmatrix}$$

deberíamos imponemos alguna de las ecuaciones del sistema, por ejemplo $x_{k+1} = x_k + y_k$, dando lugar a

$$c_{11} \left(\sqrt{2}\right)^{k+1} \cos\left(\frac{(k+1)\pi}{4}\right) + c_{21} \left(\sqrt{2}\right)^{k+1} \operatorname{sen}\left(\frac{(k+1)\pi}{4}\right) =$$
$$(c_{11} + c_{12}) \left(\sqrt{2}\right)^k \cos\left(\frac{k\pi}{4}\right) + (c_{21} + c_{22}) \left(\sqrt{2}\right)^k \operatorname{sen}\left(\frac{k\pi}{4}\right)$$

Para poder obtener una relación entre coeficientes es preciso usar las conocidas fórmulas trigonométricas

$$\cos(A+B) = \cos(A)\cos(B) - \operatorname{sen}(A)\operatorname{sen}(B)$$
$$\operatorname{sen}(A+B) = \operatorname{sen}(A)\cos(B) + \cos(A)\operatorname{sen}(B)$$

y realizar operaciones. Sin embargo, en estos casos es más conveniente obtener la relación entre los coeficientes de los vectores $\mathbf{C}_i$ razonando de modo distinto. En lugar de imponer una de las ecuaciones el sistema, podemos apoyarnos en los valores iniciales. Por ejemplo, la solución obtenida para $k = 0$ se escribe:

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix} \left(\sqrt{2}\right)^0 \cos 0 + \begin{pmatrix} c_{21} \\ c_{22} \end{pmatrix} \left(\sqrt{2}\right)^0 \operatorname{sen} 0 = \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix}$$

mientras que para $k = 1$:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix} \sqrt{2}\cos\frac{\pi}{4} + \begin{pmatrix} c_{21} \\ c_{22} \end{pmatrix} \sqrt{2}\,\text{sen}\,\frac{\pi}{4} = \begin{pmatrix} c_{11} + c_{21} \\ c_{12} + c_{22} \end{pmatrix}$$

Por otra parte, del sistema original se deduce que $x_1 = x_0 + y_0$ y $y_1 = -x_0 + y_0$, así pues

$$\begin{aligned} c_{11} + c_{21} &= c_{11} + c_{12} \\ c_{12} + c_{22} &= -c_{11} + c_{12} \end{aligned}$$

luego $c_{12} = c_{21}$, $c_{22} = -c_{11}$. La solución general queda

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{12} \end{pmatrix} \left(\sqrt{2}\right)^k \cos\left(\frac{k\pi}{4}\right) + \begin{pmatrix} c_{12} \\ -c_{11} \end{pmatrix} \left(\sqrt{2}\right)^k \text{sen}\left(\frac{k\pi}{4}\right)$$

Los cálculos realizados en el ejemplo anterior pueden generalizarse a sistemas de dimensión superior, obteniéndose el siguiente resultado.

**Teorema 7.1**

Sea $\mathbf{x}_{k+1} = A\mathbf{x}_k$ un sistema de ecuaciones lineales en diferencias de primer orden con coeficientes constantes de dimensión $n$, es decir, $A \in \mathcal{M}_n(\mathbb{R})$. Las soluciones del sistema están formadas por una combinación lineal de las siguientes soluciones:

- $\lambda^k$, si $\lambda$ es un autovalor real simple de $A$.
- $\lambda^k, k\lambda^k, \dots, k^{m-1}\lambda^k$, si $\lambda$ es un autovalor real de multiplicidad $m$ de $A$.
- $r^k\cos(k\theta)$, $r^k\,\text{sen}(k\theta)$ si $\lambda = re^{i\theta}$ es un autovalor complejo de $A$ (y por tanto, su conjugado también).
- $r^k\cos(k\theta)$, $r^k\,\text{sen}(k\theta)$, $kr^k\cos(k\theta)$, $kr^k\,\text{sen}(k\theta)$, $\dots$, $k^{m-1}r^k\cos(k\theta)$, $k^{m-1}r^k\,\text{sen}(k\theta)$, si $\lambda = re^{i\theta}$ es un autovalor complejo de multiplicidad $m$ de $A$ (y su conjugado también).

La demostración sigue las mismas ideas desarrolladas en el ejemplo 7.1, aunque los detalles resultan muy engorrosos, por lo que la omitiremos.

7 3

## ECUACIONES DE ORDEN SUPERIOR

A través del ejemplo 7.1 hemos visto que la resolución de sistemas lineales homogéneos con coeficientes constantes de primer orden se lleva a cabo mediante el cómputo de los autovalores de la matriz del sistema, a partir de los cuales obtenemos el sistema fundamental de soluciones. Veamos cómo aprovechar esta idea para abordar ecuaciones lineales con coeficientes constantes de orden superior.

Dada la ecuación en diferencias de orden $n$,

$$x_{k+n} + a_{n-1}x_{k+n-1} + \cdots + a_1 x_{k+1} + a_0 x_k = 0 \tag{7.1}$$

definimos el vector de $n$ componentes

$$\mathbf{y}_k = (x_k, x_{k+1}, \ldots, x_{k+n-1})$$

de manera que $\mathbf{y}_{k+1} = (x_{k+1}, \ldots, x_{k+n})$. Entonces la ecuación (7.1) se transforma en el sistema de primer orden

$$\mathbf{y}_{k+1} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{pmatrix} \mathbf{y}_k$$

No es difícil probar por inducción que el polinomio característico de la matriz anterior es

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

de manera que los autovalores de la matriz se deducen de las soluciones de lo que se conoce como *ecuación característica* de (7.1):

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$$

la cual puede obtenerse directamente de la expresión (7.1) sin necesidad de pasar por el sistema.

A partir de lo establecido en el Teorema 7.1 se deduce que:

(I) Si $\lambda$ es una raíz real simple de la ecuación característica, una solución vendrá dada por $\lambda^k$.

(II) Si $\lambda$ es un raíz de multiplicidad $m$ de la ecuación característica, obtendremos $m$ soluciones de la forma: $\lambda^k$, $k\lambda^k$, …, $k^{m-1}\lambda^k$.

(III) Si $\lambda = |\lambda|e^{\pm i\theta}$ es un raíz compleja (junto con su conjugada) de la ecuación característica, entonces tenemos dos soluciones del tipo $|\lambda|^k \cos(k\theta)$ y $|\lambda|^k \operatorname{sen}(k\theta)$.

(IV) Si $\lambda = |\lambda|e^{\pm i\theta}$ es un raíz compleja (junto con su conjugada) de multiplicidad $m$, entonces las $2m$ soluciones asociadas son: $|\lambda|^k \cos(k\theta)$, $k|\lambda|^k \cos(k\theta)$, ..., $k^{m-1}|\lambda|^k \cos(k\theta)$, $|\lambda|^k \operatorname{sen}(k\theta)$, $k|\lambda|^k \operatorname{sen}(k\theta)$, ..., $k^{m-1}|\lambda|^k \operatorname{sen}(k\theta)$.

**Ejemplo 7.2**

(I) Resolvamos la siguiente ecuación de orden dos

$$x_{k+2} = 2x_{k+1} - x_k$$

Consideramos la ecuación característica asociada:

$$\lambda^2 - 2\lambda + 1 = 0$$

cuyas raíces son $\lambda = 1$ doble. Entonces, la solución general de la ecuación es

$$x_k = c_1 1^k + c_2 k 1^k = c_1 + c_2 k$$

(II) Resolvamos la ecuación en diferencias

$$x_{k+4} = -x_k, \quad x_0 = x_2 = 0,\ x_1 = x_3 = 1$$

La ecuación característica es ahora: $\lambda^4 + 1 = 0$, cuyas raíces son

$$\lambda_n = e^{i\left(\frac{\pi}{4} + n\frac{\pi}{2}\right)} \text{ con } n = 0, 1, 2, 3$$

Por tanto la solución general queda

$$x_k = c_1 \cos\left(\frac{k\pi}{4}\right) + c_2 \operatorname{sen}\left(\frac{k\pi}{4}\right) + c_3 \cos\left(\frac{3k\pi}{4}\right) + c_4 \operatorname{sen}\left(\frac{3k\pi}{4}\right)$$

Para determinar la solución con los datos iniciales dados no tenemos más que sustituir:

$$x_0 = 0 \Rightarrow c_1 + c_3 = 0, \quad x_1 = 1 \Rightarrow \tfrac{\sqrt{2}}{2}(c_1 + c_2 - c_3 + c_4) = 1$$
$$x_2 = 0 \Rightarrow c_2 + c_4 = 0, \quad x_3 = 1 \Rightarrow \tfrac{\sqrt{2}}{2}(-c_1 + c_2 - c_3 - c_4) = 1$$

de donde se deduce que $c_1 = c_2 = \frac{1}{\sqrt{2}}$, $c_3 = c_4 = -\frac{1}{\sqrt{2}}$. Es decir, la solución es

$$x_k = \frac{1}{\sqrt{2}}\left(\cos\left(\frac{k\pi}{4}\right) + \operatorname{sen}\left(\frac{k\pi}{4}\right) - \cos\left(\frac{3k\pi}{4}\right) - \operatorname{sen}\left(\frac{3k\pi}{4}\right)\right)$$

### 7 3 1 Ecuaciones no homogéneas

Para resolver ecuaciones en diferencias no homogéneas haremos uso de la Proposición 7.2, es decir, primero resolveremos la ecuación homogénea y luego obtendremos una solución particular de la no homogénea. La suma de ambas nos dará la solución general de la ecuación no homogénea.

Para obtener soluciones particulares buscaremos soluciones que se asemejen a la parte no homogénea de la ecuación a resolver, introduciendo constantes indeterminadas que resolveremos imponiendo que se satisfaga la ecuación.

Consideremos la ecuación lineal de primer orden no homogénea

$$x_{k+1} = \alpha x_k + \beta \tag{7.2}$$

La solución de la ecuación homogénea la vimos al inicio de la sección 7.2, resultando $x_k = C\alpha^k$. Para obtener una solución particular de la ecuación no homogénea probaremos con una solución del tipo $x_k = A$, con $A$ constante. Sustituyendo en (7.2):

$$A = \alpha A + \beta \Rightarrow A = \frac{\beta}{1-\alpha}, \quad \text{si } \alpha \neq 1$$

Si $\alpha = 1$, podemos probar con $x_k = Ak$, que nos lleva a

$$A(k+1) = Ak + \beta \Rightarrow A = \beta$$

Así pues la solución general de la ecuación no homogénea es:

$$x_k = \begin{cases} C\alpha^k + \frac{\beta}{1-\alpha} & \text{si } \alpha \neq 1 \\ C\alpha^k + \beta k & \text{si } \alpha = 1 \end{cases}$$

La constante $C$ quedará determinada con la condición inicial.

Veamos algunos ejemplos:

**Ejemplo 7.3**

(i) Resolver la ecuación de primer orden

$$x_{k+1} = 5x_k + 3^k, \quad x_0 = 1$$

Al igual que antes, resolvemos en primer lugar la ecuación homogénea $x_{k+1} = 5x_k$ cuya solución general es $x_k = C5^k$. Para obtener una solución

particular de la ecuación no homogénea probaremos con una sucesión del tipo $x_k = A3^k$. Sustituyendo en la ecuación

$$A3^{k+1} = 5A3^k + 3^k \Rightarrow A = -\frac{1}{2}$$

La solución general es $x_k = -\frac{1}{2}3^k + C5^k$. Para obtener la solución con condición inicial $x_0 = 1$ bastará sustituir en la expresión anterior para obtener que $C = \frac{3}{2}$ y por tanto

$$x_k = -\frac{1}{2}3^k + \frac{3}{2}5^k$$

(II) Resolvamos la ecuación

$$x_{k+2} + x_{k+1} + x_k = k^2 + k + 1, \quad x_0 = x_1 = 0$$

Para resolver la ecuación homogénea observamos que la ecuación característica es

$$\lambda^2 + \lambda + 1 = 0$$

cuyas raíces son $\lambda = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i = e^{\pm\frac{2\pi}{3}i}$. Luego la solución general de la ecuación homogénea es

$$x_k = c_1 \cos\left(\frac{2k\pi}{3}\right) + c_2 \operatorname{sen}\left(\frac{2k\pi}{3}\right)$$

Por otra parte, buscaremos una solución particular de la forma $x_k = Ak^2 + Bk + C$. Sustituyendo,

$$A(k+2)^2+B(k+2)+C+A(k+1)^2+B(k+1)+C+Ak^2+Bk+C = k^2+k+1$$

se obtiene

$$A = \frac{1}{3}, \quad B = -\frac{1}{3}, \quad C = \frac{1}{9}$$

obteniéndose así la solución general de la ecuación dada:

$$x_k = c_1 \cos\left(\frac{2k\pi}{3}\right) + c_2 \operatorname{sen}\left(\frac{2k\pi}{3}\right) + \frac{1}{3}k^2 - \frac{1}{3}k + \frac{1}{9}$$

Sustituyendo esta expresión para $x_0 = x_1 = 0$ se obtiene que $c_1 = -\frac{1}{9}$ y $c_2 = -\frac{1}{3\sqrt{3}}$.

(III) Resolvamos la ecuación en diferencias

$$x_{k+4} + x_k = 2k + 1$$

La solución general de la ecuación homogénea la hemos calculado en (II) del ejemplo 7.2:

$$x_k = c_1 \cos\left(\frac{k\pi}{4}\right) + c_2 \operatorname{sen}\left(\frac{k\pi}{4}\right) + c_3 \cos\left(\frac{3k\pi}{4}\right) + c_4 \operatorname{sen}\left(\frac{3k\pi}{4}\right)$$

Por otro lado, buscaremos una solución particular de la ecuación no homogénea que sea de la forma $x_k = Ak + B$, obteniéndose

$$A(k+4) + B + Ak + B = 2k + 1 \Rightarrow A = 1,\ B = -\frac{3}{2}$$

y por tanto, la solución general de la ecuación es

$$x_k = c_1 \cos\left(\frac{k\pi}{4}\right) + c_2 \operatorname{sen}\left(\frac{k\pi}{4}\right) + c_3 \cos\left(\frac{3k\pi}{4}\right) + c_4 \operatorname{sen}\left(\frac{3k\pi}{4}\right) + k - \frac{3}{2}$$

(IV) Resolver el sistema de ecuaciones en diferencias

$$\left\{\begin{array}{rcl} x_{k+1} & = & y_k + k \\ y_{k+1} & = & 9x_k \end{array}\right. \qquad x_0 = 0,\ y_0 = 1$$

Procedemos a calcular la solución general del sistema homogéneo asociado

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 9 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

cuyos autovalores son 3 y $-3$, y los autovectores $(1, 3)$ y $(1, -3)$, respectivamente; de modo que la solución general vendrá dada por

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_{11} \\ 3c_{11} \end{pmatrix} 3^k + \begin{pmatrix} c_{21} \\ -3c_{21} \end{pmatrix} (-3)^k$$

Para buscar una solución particular procedemos de forma similar a los ejemplos anteriores, es decir, buscamos un vector de la forma

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} A_1 k + B_1 \\ A_2 k + B_2 \end{pmatrix}$$

Imponiendo que esta expresión sea solución de la ecuación en diferencias:

$$\begin{aligned} A_1(k+1) + B_1 &= A_2 k + B_2 + k \\ A_2(k+1) + B_2 &= 9(A_1 k + B_1) \end{aligned}$$

se obtiene el sistema

$$\begin{cases} A_1 - A_2 &= 1 \\ A_1 + B_1 - B_2 &= 0 \\ A_2 - 9A_1 &= 0 \\ A_2 + B_2 - 9B_1 &= 0 \end{cases}$$

cuya solución es $A_1 = -\frac{1}{8}$, $B_1 = -\frac{5}{32}$, $A_2 = -\frac{9}{8}$ y $B_2 = -\frac{9}{32}$. La solución general de la ecuación no homogénea será

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} c_{11} \\ 3c_{11} \end{pmatrix} 3^k + \begin{pmatrix} c_{21} \\ -3c_{21} \end{pmatrix} (-3)^k + \begin{pmatrix} -\frac{1}{8}k - \frac{5}{32} \\ -\frac{9}{8}k - \frac{9}{32} \end{pmatrix}$$

Imponiendo ahora la condición inicial $x_0 = 0$, $y_0 = 1$ se tiene

$$\begin{cases} c_{11} + c_{21} - \frac{5}{32} = 0 \\ 3c_{11} - 3c_{21} - \frac{9}{32} = 1 \end{cases} \Rightarrow c_{11} = \frac{7}{24},\ c_{21} = -\frac{13}{96}$$

7 4

## CADENAS DE MARKOV

Una de las aplicaciones típicas de los sistemas dinámicos discretos aparece cuando los sistemas de ecuaciones en diferencias tratan con cantidades porcentuales que pasan de un estado a otro sin variar en su conjunto. Veámoslo en el siguiente ejemplo.

**Ejemplo 7.4**

Supongamos que los $N$ habitantes de una cierta ciudad realizan sus compras en uno de los tres supermercados existentes $X$, $Y$ y $Z$. Considerando que las grandes compras se realizan una vez por semana, y que los consumidores comparan calidad, precio, comodidad, etc., algunos deciden cambiar de supermercado, mientras que otros permanecen fieles. Pongamos que el porcentaje de clientes que al cabo de una semana deciden permanecer o cambiar de supermercado está dado en la siguiente tabla:

| | $X$ | $Y$ | $Z$ |
|---|---|---|---|
| $X$ | .80 | .20 | .10 |
| $Y$ | .10 | .70 | .30 |
| $Z$ | .10 | .10 | .60 |

donde el porcentaje $p$ que aparece en la fila $i$ y la columna $j$ significa que si en una semana hay $m$ clientes que compran en el supermercado $j$, la semana siguiente habrá $pm$ clientes que compran en el supermercado $i$. Suponiendo que el número de clientes es constante, se trata de averiguar cuál será el mercado que atraiga al mayor número de compradores.

Si denotamos por $x_k$, $y_k$ y $z_k$ al número de clientes que compran en los supermercados $X$, $Y$ y $Z$ en la semana $k$, respectivamente, es fácil darse cuenta que

$$\begin{aligned} x_{k+1} &= 0.80x_k + 0.20y_k + 0.1z_k \\ y_{k+1} &= 0.10x_k + 0.70y_k + 0.3z_k \\ z_{k+1} &= 0.10x_k + 0.10y_k + 0.6z_k \end{aligned} \tag{7.3}$$

es decir, tenemos un sistema de ecuaciones en diferencias lineal homogéneo.

El anterior es un típico ejemplo de lo que se conoce como una *cadena* o *sistema de Markov*.[2]

**Definición 7.4**

Diremos que una matriz real cuadrada es *de Markov* si sus elementos son no negativos y sus columnas suman uno.

La interpretación típica en términos de probabilidad de una matriz de Markov es la de un sistema que consta de $n$ estados posibles $E_1, \dots, E_n$ en la que la probabilidad de pasar del estado $E_j$ al estado $E_i$ viene dada por el elemento $a_{ij}$.

Veamos algunas propiedades interesantes de estos sistemas.

[2] Introducidas por el matemático ruso Andréi Andréyevich Markov en 1907.

**Proposición 7.3**

Si $A$ es una matriz de Markov entonces:

(I) Todos sus autovalores $\lambda$ verifican que $|\lambda| \leq 1$.

(II) $\lambda = 1$ es autovalor de $A$.

*Demostración:*

(I) Procederemos por reducción al absurdo. Supongamos que $|\lambda| > 1$ es un autovalor de $A$ asociado a un autovector $\mathbf{v}$, que podemos suponer, sin pérdida de generalidad, que satisface que

$$\sum_{i=1}^{n} |v_i| = 1$$

Entonces

$$1 < |\lambda| = |\lambda| \sum_{i=1}^{n} |v_i| = \sum_{i=1}^{n} |(\lambda \mathbf{v})_i| = \sum_{i=1}^{n} |(A\mathbf{v})_i| = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} v_j \right|$$

$$\leq \sum_{i=1}^{n} \left( \sum_{j=1}^{n} |a_{ij} v_j| \right) = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} a_{ij} \right) |v_j| = \sum_{j=1}^{n} |v_j| = 1$$

pues las columnas de $A$ suman uno y los elementos $a_{ij}$ son positivos.

(II) Puesto que las columnas de $A$ suman uno, si $\mathbf{e} = (1, \ldots, 1)$ se tiene que $\mathbf{e}^T A = \mathbf{e}^T$. Trasponiendo esta expresión: $A^T \mathbf{e} = \mathbf{e}$, por tanto $\mathbf{e}$ es un autovector de $A^T$ asociado al autovalor 1. Puesto que una matriz y su traspuesta tienen los mismos autovalores, se tiene el resultado.

■

**Definición 7.5**

Un vector $\mathbf{v}$ se dice *vector de probabilidad* si sus componentes son todas no negativas y suman uno.

En la demostración de (I) de la Proposición 7.3 se demuestra implícitamente el siguiente resultado:

**Proposición 7.4**

Si $A$ es una matriz de Markov y $\mathbf{v}$ un vector de probabilidad, entonces $A\mathbf{v}$ también es un vector de probabilidad.

El significado de este resultado es el siguiente: en un proceso de Markov, "no se gana ni se pierde nada", es decir, la cantidad total que hay al inicio permanece constante durante todo el proceso.

**Definición 7.6**

Se denomina *vector estacionario* o de *equilibrio* de una matriz de Markov a todo vector de probabilidad $\mathbf{v}$ tal que $A\mathbf{v} = \mathbf{v}$.

En particular, los vectores de equilibrio de una matriz de Markov son los autovectores asociados al autovalor 1.

**Ejemplo 7.5 (continuación del ejemplo 7.4)**

El sistema (7.3) se representa en forma matricial por $\mathbf{x}_{k+1} = A\mathbf{x}_k$, donde

$$\mathbf{x}_k = \begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} \qquad A = \begin{pmatrix} 0.8 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}$$

de modo que $\mathbf{x}_k = A^k\mathbf{x}_0$. Supongamos que el dato inicial es $(x_0, y_0, z_0) = (0.2, 0.3, 0.5)$, es decir, que inicialmente, hay un 20 % de la población que va al supermercado $X$, un 30 % va al $Y$ y un 50 % va al $Z$. Tratemos de averiguar la tendencia del mercado, es decir, qué porcentaje de población acudirá a cada uno de los supermercados pasado un cierto tiempo.

Teniendo en cuenta que los autovalores de $A$ son 1, 0.6 y 0.5 se tendrá que la solución general del sistema vendrá dada por

$$\mathbf{x}_k = \mathbf{C}_1 + \mathbf{C}_2(0.6)^k + \mathbf{C}_3(0.5)^k$$

Dado que esto es un sistema tridimensional, la solución general deberá depender de tres constantes, y no de nueve como indica la aparición de los vectores $\mathbf{C}_i$. Podríamos proceder de forma similar a los ejemplos de sistemas de ecuaciones en diferencias que hemos tratado, imponiendo en este caso dos de las ecuaciones

para reducir el número de constantes o usando los estados iniciales, pero esto es bastante tedioso.

En este caso, razonaremos de otro modo. Si calculamos los autovectores asociados a cada autovalor obtenemos los vectores $(9,7,4)$, $(1,-1,0)$ y $(1,-2,1)$ asociados a 1, 0.6 y 0.5 respectivamente. Calculemos ahora $(\alpha_1,\alpha_2,\alpha_3)$ tales que

$$(0.2,0.3,0.5)=\alpha_1(9,7,4)+\alpha_2(1,-1,0)+\alpha_3(1,-2,1)\Rightarrow\begin{cases}\alpha_1=0.05\\ \alpha_2=-0.55\\ \alpha_3=0.3\end{cases}$$

Entonces, el estado límite

$$\begin{aligned}\mathbf{x}_\infty=\lim_{k\to\infty}A^k\mathbf{x}_0&=\lim_{k\to\infty}A^k\left[0.05\begin{pmatrix}9\\7\\4\end{pmatrix}-0.55\begin{pmatrix}1\\-1\\0\end{pmatrix}+0.3\begin{pmatrix}1\\-2\\1\end{pmatrix}\right]\\&=\lim_{k\to\infty}\left[0.05A^k\begin{pmatrix}9\\7\\4\end{pmatrix}-0.55A^k\begin{pmatrix}1\\-1\\0\end{pmatrix}+0.3A^k\begin{pmatrix}1\\-2\\1\end{pmatrix}\right]\\&=\lim_{k\to\infty}\left[0.05\cdot 1^k\begin{pmatrix}9\\7\\4\end{pmatrix}-0.55\cdot(0.6)^k\begin{pmatrix}1\\-1\\0\end{pmatrix}+0.3\cdot(0.5)^k\begin{pmatrix}1\\-2\\1\end{pmatrix}\right]\\&=\begin{pmatrix}0.45\\0.35\\0.2\end{pmatrix}\end{aligned}$$

**Nota 7.1**

En el ejemplo anterior hemos usado que si $\mathbf{v}$ es un autovector de $A$ asociado a un autovalor $\lambda$, entonces $\mathbf{v}$ es un autovector de $A^k$ asociado al autovalor $\lambda^k$,

es decir,

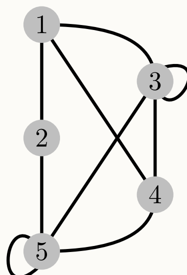
$$A\mathbf{v} = \lambda\mathbf{v} \Rightarrow A^k\mathbf{v} = \lambda^k\mathbf{v}$$

Por otro lado, los cálculos realizados son válidos en cualquier situación, es decir, aquí no se ha usado que la matriz es de Markov.

Más aun, en este caso concreto debemos observar que nos hubiera sobrado con calcular el autovector asociado al autovalor 1, pues en el estado límite, los autovectores asociados a autovalores menores que uno desaparecen, por lo que el estado límite es un vector estacionario.

**Ejemplo 7.6**

Un ratón se encuentra en un laberinto con 5 compartimentos que están conectados entre sí a través de túneles tal y como se representa en la siguiente figura.




Se supone que cada vez que el ratón deja un compartimento y elige un túnel que lo conecta, lo hace con igual probabilidad, y que en cada etapa, el ratón siempre elige un túnel para moverse. Encontrar en qué compartimento pasará el ratón la mayor parte del tiempo, tras un período suficientemente largo.

Según la figura, la probabilidad de que el ratón se mueva de un compartimento a otro en una determinada etapa vendrá dada por la siguiente matriz:

$$A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{5} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{5} \\ \frac{1}{3} & 0 & \frac{2}{5} & \frac{1}{3} & \frac{1}{5} \\ \frac{1}{3} & 0 & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & \frac{1}{2} & \frac{1}{5} & \frac{1}{3} & \frac{2}{5} \end{pmatrix}$$

donde el elemento de la fila $i$ columna $j$ representa la probabilidad de que el

ratón vaya desde el compartimento $j$ al $i$.

Denotando por $x_i^{(k)}$, a la probabilidad de que en la etapa $k$ el ratón esté en el compartimento $i$, el estado del sistema vendrá dado por $\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$. Suponiendo que el ratón se encuentra en el estado inicial en cualquiera de los compartimentos con igual probabilidad, es decir, el estado inicial $\mathbf{x}^{(0)} = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ veamos cuál es el estado límite.

Puesto que se trata de un proceso de Markov, todos los estados asociados a autovalores menores que uno desparecerá en el límite, por lo que solo hemos de preocuparnos por los estados estacionarios, es decir, los autovectores asociados al autovalor 1.

Si calculamos $\ker(A - I)$ se obtiene como autovector independiente $\mathbf{v} = (3, 2, 5, 3, 5)$, de modo que $\mathbf{x}^{(k)} = A^k\mathbf{x}^{(0)} = c\mathbf{v}$, donde $c$ es una constante que determinamos observando que el resultado tiene que ser un vector de probabilidad (según la Proposicion 7.4). Así, la suma de sus elementos debe ser uno y por tanto $c = \frac{1}{18}$. El estado límite corresponde al vector $(\frac{1}{6}, \frac{1}{9}, \frac{5}{18}, \frac{1}{6}, \frac{5}{18})$, lo que significa que pasará el mayor tiempo en los compartimentos 3 y 5.

## 7 5 APLICACIÓN: MODELOS BIOLÓGICOS

Un ejemplo típico de aplicación de los sistemas dinámicos en Biología aparece en los modelos de evolución de dos especies competidoras que interactúan unas con otras, como los modelos *depredador-presa*.

De manera simplificada se suele suponer que el crecimiento de una determinada especie de presas (altamente reproductoras, como por ejemplo, los conejos) en ausencia de depredadores crece según la ecuación

$$x_{k+1} = ax_k, \quad a > 1$$

donde $x_k$ denota el número de presas en un instante temporal $k$, y $a$ es un parámetro que estima la tasa de reproducción de la especie. Nótese que al ser $a > 1$, el número de individuos de este especie crece en cada etapa de tiempo. Por otra parte, el crecimiento de los depredadores en ausencia de presas (y por tanto sin alimento) es negativo, modelándose por

$$y_{k+1} = dy_k, \quad 0 < d < 1$$

La interacción de ambas especies nos lleva a un sistema como el siguiente

$$\begin{aligned} x_{k+1} &= ax_k - by_k \\ y_{k+1} &= dy_k + cx_k \end{aligned}$$

en el que los parámetros $b$ y $c$ miden la influencia que cada especie tiene en el crecimiento de población de la otra. Así, $b > 0$ indica que la presencia de depredadores disminuye la población de presas, mientras que $c > 0$ significa que la cantidad de presas (es decir, de alimento) incrementa el número de depredadores. El sistema anterior admite una representación matricial del tipo

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} a & -b \\ c & d \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

que no es más que un sistema de ecuaciones en diferencias de orden uno. Veámoslo con más detenimiento a través de un ejemplo concreto:

Consideremos una población de zorros en una región de un bosque cuyo alimento preferido son los conejos. Denotemos por $Z_k$ y $C_k$ al número de individuos (en miles) de cada especie medidos en un período de tiempo $k$, y supongamos que la evolución de las dos poblaciones obedece a las ecuaciones siguientes:

$$\begin{aligned} Z_{k+1} &= 0.5Z_k + 0.4C_k \\ C_{k+1} &= -pZ_k + 1.1C_k \end{aligned}$$

donde $p$ es un parámetro positivo a especificar. Como vemos, en ausencia de conejos, la población de zorros se divide por la mitad en cada etapa temporal, mientras que en ausencia de zorros, la población de conejos aumenta un 10 % en cada etapa. Sin embargo, la aparición de conejos incrementa la población de zorros, mientras que la existencia de zorros hace que la población de conejos disminuya.

Imaginemos en primer lugar que el parámetro $p = 0.2$, y veamos cuál es la evolución de las poblaciones a largo plazo. En este caso el sistema dinámico queda:

$$\begin{pmatrix} Z_{k+1} \\ C_{k+1} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.4 \\ -0.2 & 1.1 \end{pmatrix} \begin{pmatrix} Z_k \\ C_k \end{pmatrix}$$

Vamos a usar Python para realizar los cálculos de autovalores y autovectores:

```
>>> from sympy import Matrix
>>> a=Matrix([[0.5,0.4],[-0.2,1.1]])
>>> a.eigenvects()
[(0.700000000000000, 1, [[2.0]
[  1]]), (0.900000000000000, 1, [[1.0]
[  1]])]
```

Esto es, los autovalores y autovectores asociados son:

$$0.7 \to (2, 1) \qquad 0.9 \to (1, 1)$$

¿Qué significado tiene el resultado? Teniendo en cuenta que los autovalores son 0.7 y 0.9, el sistema tenderá a algo del estilo

$$\begin{pmatrix} Z_k \\ C_k \end{pmatrix} = \mathbf{A}_1(0.7)^k + \mathbf{A}_2(0.9)^k \xrightarrow[k\to\infty]{} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Es decir, para el valor de $p = 0.2$, las dos especies se extinguirán. Es fácil intuir el por qué: la voracidad de los zorros hace que la población de conejos se extinga y posteriormente, al quedarse sin alimento, desaparecen también los zorros.

La pregunta ahora sería, cuánto debe valer $p$ para que se mantengan constantes ambas poblaciones. Si los autovalores son menores que uno, acabamos de ver que ambas poblaciones están condenadas a la extinción; si ambos son mayores que uno, las poblaciones se disparan, es decir ambas poblaciones crecen indefinidamente (lo que desde un punto de vista realista no tiene sentido, pues los recursos son limitados, pero este es un efecto que no estamos considerando en el sistema). Por otro lado, si un autovalor es mayor que uno y el otro menor que uno, en función del dato inicial se tendría que las poblaciones pueden crecer o extinguirse. Se observa que la única forma de llegar al equilibrio es que alguno de los autovalores sea igual a uno.

```
1  >>> from sympy import Matrix,symbols,eye,solve
2  >>> p=symbols('p')
3  >>> a=Matrix([[0.5,0.4],[-p,1.1]])
4  >>> b=a-eye(2)
5  >>> b.det()
6  -0.05 + 0.4*p
7  >>> solve(_)
8  [0.125000000000000]
9  >>> a.subs(p,0.125).eigenvals()
10 {0.600000000000000: 1, 1.00000000000000: 1}
```

luego

$$\begin{vmatrix} 0.5-1 & 0.4 \\ -p & 1.1-1 \end{vmatrix} = -0.04 + 0.4p = 0 \Rightarrow p = 0.125$$

Es decir, para $p = 0.125$ los autovalores son 1 y 0.6.

Obsérvese el uso del guión bajo para reutilizar la última salida en Python (línea 7) y el módulo `subs` (línea 9) para sustituir el valor de $p$ en la matriz y luego calcular los autovalores.

Para el valor de $p = 0.125$ aseguramos que uno de los autovalores es 1 y el otro es menor que uno, de forma que el vector asociado al autovalor 1 nos dará estados en los que la población se mantiene constante. Si intentamos obtener el autovector asociado al autovalor 1:

```
11 >>> a.subs(p,0.125).eigenvects()
12 [(0.600000000000000, 1, [[4.0]
13 [  1]]), (1.00000000000000, 1, [])]
```

nos encontramos con que Python no es capaz de proporcionarlo, posiblemente por algún error de redondeo.[3] En este caso hemos de calcularlo a mano, resultando el vector $(4,5)$. Es decir, si el estado inicial es un múltiplo de este vector, entonces ambas poblaciones se mantienen constantes.

7 6

## EJERCICIOS

### Ejercicios de repaso

**E.1** Resolver las siguientes ecuaciones en diferencias:

$$\begin{aligned}
&(a)\ x_{n+2} - x_{n+1} - 2x_n = 0,\ x_1 = 0,\ x_2 = 5\\
&(b)\ x_{n+2} - 5x_{n+1} + 6x_n = 0,\ x_1 = 1,\ x_2 = 2\\
&(c)\ x_{n+3} - 5x_{n+2} + 3x_{n+1} + 9x_n = 0\\
&(d)\ x_{n+4} + 2x_{n+2} + x_n = 0
\end{aligned}$$

**E.2** Se dan las sucesiones recurrentes

$$\begin{aligned}
u_n &= 3u_{n-1} + 3v_{n-1}\\
v_n &= 5u_{n-1} + v_{n-1}
\end{aligned}$$

con $u_0 = 1$ y $v_0 = 1$. Hallar $u_n$ y $v_n$ en función de $n$.

**E.3** Sea la sucesión (de Fibonacci) $0, 1, 1, 2, 3, 5, 8, 13, \dots$ i.e. $x_{n+2} = x_{n+1} + x_n$. Calcula el término $x_n$ en función de $n$.

**E.4** Calcular, si es posible, el estado límite para el sistema cuya matriz es

$$\begin{pmatrix} 0.9 & 0.15\\ 0.1 & 0.85 \end{pmatrix}$$

**E.5** Estudiar la dinámica de un punto cuyas coordenadas en el instante $n$ vienen dadas por las ecuaciones

$$\begin{pmatrix} x_n\\ y_n\\ z_n \end{pmatrix} = \begin{pmatrix} \frac{5}{2} & 3 & 0\\ -\frac{3}{2} & -2 & 0\\ -6 & -6 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_{n-1}\\ y_{n-1}\\ z_{n-1} \end{pmatrix}$$

### Problemas

**E.6** Sea la sucesión $0, 1, 2, 5, 12, 29, \dots$, i.e. $x_n = 2x_{n-1} + x_{n-2}$. Demostrar que

$$x_n = \frac{\left(1+\sqrt{2}\right)^n}{2\sqrt{2}} - \frac{\left(1-\sqrt{2}\right)^n}{2\sqrt{2}}$$

[3] Este comportamiento, aunque anómalo, puede ocurrir en determinadas ocasiones debido a las limitaciones o posible errores del software.

y usar el hecho de que $\frac{x_{n+1}}{x_n} \approx 1 + \sqrt{2}$ para aproximar $\sqrt{2}$.

**E.7** Se sabe que la posición de una partícula en el plano en el instante $t \geq 0$ esta dada por

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}$$

Estudiar la dinámica de dicha partícula según sea la posición inicial $(x_0, y_0)$.

**E.8** Los coches españoles sin catalizador se dividen en tres grupos: Los que cada año han de pasar la inspección técnica de vehículos (la ITV), los que aún son suficientemente nuevos como para no tener que pasarla y los que causan baja (van a la chatarra). Se sabe que el 2 % de los coches nuevos tienen que pasar la ITV al año siguiente y que el 3 % de los que ya la pasaban causan baja. Se supone que ya no se fabrican coches sin catalizador y que por tanto el número de coches sin catalizador es fijo. Si en el año 95 hay 3000 coches que deben pasar revisión y 15000 que todavía no, calcula cuántos años han de pasar para que se reduzca a la mitad la cantidad de coches que no tienen que pasar la ITV.

**E.9** Un estudio realizado sobre la comunidad de ingenieros industriales revela el hecho siguiente: El 90 % de los hijos de padres ingenieros industriales cursan estudios de ingeniería industrial y sólo el 20 % de los hijos de padres no ingenieros industriales cursan esa carrera. ¿Cuál será el porcentaje de estudiantes que cursarán la carrera de ingeniería industrial después de muchas generaciones? (se supondrá un hijo como descendencia por cada familia).

**E.10** El ascensor de un edificio con bajo y dos pisos realiza viajes de uno a otro piso. Se sabe que la mitad de los viajes que parten del bajo se dirigen a cada uno de los otros dos pisos, mientras que si un viaje comienza en el primer piso, sólo el 25 % de las veces finaliza en el segundo. Por último, si un trayecto comienza en el segundo piso, siempre finaliza en el bajo. ¿Cuál es la probabilidad de que, a largo plazo, el ascensor se encuentre en cada uno de los tres pisos?

**E.11** Las familias de cierto país se clasifican según residan en áreas rurales, urbanas o suburbanas. Los estudios de movilidad demográfica estiman que, en promedio, en el curso de un año, el 15 % de las familias urbanas cambia de residencia y se traslada a un área suburbana, y el 5 % a un área rural; mientras que el 6 % de las familias residentes en áreas suburbanas se traslada a áreas urbanas, y el 4 % a áreas rurales, y finalmente el 4 % de las familias rurales migra a las áreas urbanas y el 6 % a las suburbanas. Supongamos que en el momento actual el 40 % de las familias del país viven en áreas urbanas, el 35 % en suburbanas y el 25 % en rurales. ¿Qué distribución de población es de prever en el futuro si las tendencias no cambian?

## Ejercicios teóricos

*E.12 Se considera un sistema dinámico $\mathbf{x}_{k+1} = A\mathbf{x}_k$ con matriz $A$ diagonalizable. Se define el *radio espectral* de $A$, y se denota por $\rho(A)$ como

$$\rho(A) = \max_{1\leq i\leq n} \{|\lambda_1|, |\lambda_2|, \dots, |\lambda_n|\}$$

con $\lambda_1, \dots, \lambda_n$ los autovalores de $A$. Probar:

(a) Si $\rho(A) < 1$ entonces $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{0}$, $\forall \mathbf{x}_0$.

(b) Si $\rho(A) = 1$, existe algún $\mathbf{x}_0$ tal que $\lim_{k\to\infty} \mathbf{x}_k \neq \mathbf{0}$.

E.13 Probar que si $\lambda$ es un autovalor asociado a un autovector $\mathbf{v}$ de una matriz $A \in \mathcal{M}_n(\mathbb{K})$, entonces $\lambda^k$ es autovalor asociado a $\mathbf{v}$ de $A^k$.

E.14 Sea $A$ una matriz de Markov de orden $2 \times 2$ distinta de la identidad.

(a) Probar que existen números $a$ y $b$, con $0 \leq a, b \leq 1$ tales que

$$A = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}$$

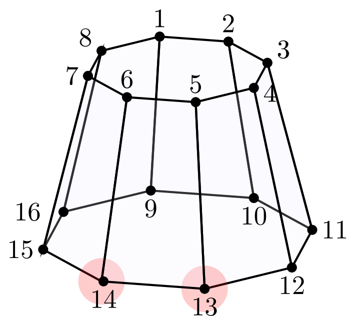
(b) Compruébese que los vectores $(b, a)$ y $(1, -1)$ son autovectores de $A$.

(c) Hallar la forma canónica de $A$ y obtener una expresión de $A^k$.

## Ejercicios adicionales

*E.15

Una hormiga se encuentra situada en el vértice 1 de la pirámide truncada que aparece en la figura adjunta.




Cada minuto que pasa, la hormiga se desplaza exclusivamente por las aristas de la pirámide hasta un vértice adyacente, eligiendo con igual probabilidad la arista por la que se mueve. Así, por ejemplo, en el primer minuto la hormiga se moverá o al vértice 2, o al vértice 8, o al vértice 9, con probabilidad $\frac{1}{3}$ en todos los casos.

Rociamos con un spray insecticida los vértices 13 y 14 de la pirámide, de manera que si la hormiga llega a cualquiera de estos dos vértices morirá de inmediato. ¿Cuál es la probabilidad de que la hormiga siga viva al cabo de 10 minutos? ¿Y tras una hora? Hay alguna posibilidad de que la hormiga sobreviva al cabo del tiempo?

Sugerencia: realizar los cálculos con la ayuda de Python.

# 8 Espacio vectorial euclídeo

## 8.1 PRODUCTO ESCALAR

En este tema vamos a introducir una nueva operación en un espacio vectorial, el *producto escalar*, que nos va a dar acceso a nuevas e interesantes posibilidades. Entre ellas, el producto escalar nos va a permitir introducir el concepto de longitud en un espacio vectorial, y gracias a éste, podremos calcular *distancias* entre vectores e introducir una noción de proximidad que tiene importantes aplicaciones.

La introducción de esta nueva operación dota a los espacios vectoriales de una estructura nueva, denominada *espacio euclídeo*, cuya definición damos a continuación.

**Definición 8.1**

Sea $E$ un espacio vectorial real. Una aplicación

$$\begin{array}{rccc} \langle\cdot,\cdot\rangle : & E\times E & \longrightarrow & \mathbb{R} \\ & (\mathbf{x},\mathbf{y}) & \longrightarrow & \langle\mathbf{x},\mathbf{y}\rangle \end{array}$$

se dice que es un *producto escalar* si satisface las siguientes propiedades:

(I) $\langle\mathbf{x},\mathbf{y}\rangle = \langle\mathbf{y},\mathbf{x}\rangle$, $\forall\mathbf{x},\mathbf{y}\in E$.

(II) $\langle\mathbf{x}+\mathbf{y},\mathbf{z}\rangle = \langle\mathbf{x},\mathbf{z}\rangle + \langle\mathbf{y},\mathbf{z}\rangle$, $\forall\mathbf{x},\mathbf{y},\mathbf{z}\in E$.

(III) $\langle\alpha\mathbf{x},\mathbf{y}\rangle = \alpha\langle\mathbf{x},\mathbf{y}\rangle$, $\forall\mathbf{x},\mathbf{y}\in E$, $\forall\alpha\in\mathbb{R}$.

(IV) $\langle\mathbf{x},\mathbf{x}\rangle \geq 0$, $\forall\mathbf{x}\in E$. Además $\langle\mathbf{x},\mathbf{x}\rangle = 0 \Leftrightarrow \mathbf{x}=\mathbf{0}$.

Si en un espacio vectorial $E$ hay definido un producto escalar se dirá que $E$ es un *espacio euclídeo* (e.e.).

La propiedad (I) se denomina de *simetría*, (II) y (III) son propiedades de *linealidad*, mientras que la propiedad (IV) se conoce como *positividad*. Nótese

también que de (I)—(III) se deduce

$$\left.\begin{array}{c}\langle \mathbf{x}, \mathbf{y}+\mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{x}, \mathbf{z}\rangle \\ \langle \mathbf{x}, \alpha\mathbf{y}\rangle = \alpha\langle \mathbf{x}, \mathbf{y}\rangle\end{array}\right. \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in E, \ \ \forall \alpha \in \mathbb{R}$$

**Nota 8.1**

Si $E$ es un espacio vectorial complejo, también se puede dar una definición similar para lo que se conoce genéricamente como *producto interno*. En este caso se trata de un aplicación

$$\begin{array}{rccc}\langle\cdot,\cdot\rangle: & E\times E & \longrightarrow & \mathbb{C} \\ & (\mathbf{x}, \mathbf{y}) & \longrightarrow & \langle \mathbf{x}, \mathbf{y}\rangle\end{array}$$

que satisface (II)–(IV), y en la que (I) se sustituye por

$$\langle \mathbf{x}, \mathbf{y}\rangle = \overline{\langle \mathbf{y}, \mathbf{x}\rangle}, \quad \forall \mathbf{x}, \mathbf{y} \in E$$

En tal caso diremos que $E$ es un *espacio unitario*.

**Ejemplo 8.1**

(I) En $\mathbb{R}^n$, la aplicación

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n} x_i y_i$$

es un producto escalar, que será el producto escalar habitual que usaremos en este espacio. La notación habitual para el mismo es $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y}$, o en forma matricial, $\mathbf{x}^T\mathbf{y}$.

(II) En $\mathbb{R}^2$,

$$\langle \mathbf{x}, \mathbf{y}\rangle = \begin{pmatrix} x_1 & x_2\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}y_1\\ y_2\end{pmatrix}$$

es un producto escalar.

(III) $\mathcal{C}([a,b])$ es un espacio euclídeo con el producto

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$

(IV) En $\mathbb{R}^2$, $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1$ no es un producto escalar puesto que no verifica la propiedad de positividad. Obsérvese que, por ejemplo, $\mathbf{x} = (0,1)$ no la satisface, pues

$$\langle (0,1), (0,1) \rangle = 0, \text{ pero } (0,1) \text{ no es el vector nulo.}$$

**Nota 8.2**

En $\mathbb{C}^n$, el producto interno habitual es $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$.

En todo lo que sigue consideraremos que $E$ es un espacio vectorial real.

**Definición 8.2**

Sea $E$ un e.e. Se denomina *norma* o *longitud* de un vector $\mathbf{x}$ a $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Se notará por $\|\mathbf{x}\|$.

La norma es un concepto fundamental, pues nos permite cuantificar o medir el tamaño de un vector. Si un vector tiene norma próxima a cero, será un vector pequeño, y viceversa. Gracias a esto, dos vectores son próximos si la norma de su diferencia es pequeña (véase la sección 3.3.2).

El siguiente resultado es una desigualdad notable que relaciona las normas de dos vectores con su producto escalar.

**Proposición 8.1 (Desigualdad de Schwarz[1])**

Si $\mathbf{x}$, $\mathbf{y} \in E$, un e.e., entonces

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \leq 1$$

[1] También denominada desigualdad de Cauchy-Schwarz, aunque la formulación y prueba que mostramos aquí se debe al alemán Hermann Klaus Hugo Weyl.

*Demostración:*
Sea $\lambda \in \mathbb{R}$. La propiedad de positividad del producto escalar implica que

$$\langle \lambda \mathbf{x} - \mathbf{y}, \lambda \mathbf{x} - \mathbf{y} \rangle \geq 0, \quad \forall \mathbf{x}, \mathbf{y} \in E, \ \forall \lambda \in \mathbb{R}$$

Desarrollando la expresión anterior haciendo uso de las propiedades de linealidad del producto escalar se tiene que

$$\lambda^2 \langle \mathbf{x}, \mathbf{x} \rangle - 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \geq 0, \ \forall \lambda \in \mathbb{R}$$

Esta última expresión puede ser vista como un polinomio de segundo grado en $\lambda$ que, dado que es positivo, tendrá a lo más una raíz. Es decir, el discriminante de la ecuación de segundo grado correspondiente debe ser menor o igual que cero. De este modo,

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0 \Rightarrow \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2} \leq 1$$

de donde se sigue el resultado.

■

Como consecuencia de la desigualdad de Schwarz obtenemos otras desigualdades relevantes en la literatura matemática.

**Desigualdad de Cauchy**:

$$\sum_{i=1}^{n} x_i y_i \leq \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \left( \sum_{i=1}^{n} y_i^2 \right)^{1/2}$$

(que se obtiene al aplicar la desigualdad de Schwarz al producto escalar (I) del ejemplo 8.1).

**Desigualdad de Bunyakowski**:

$$\int_a^b f(x) g(x) \, dx \leq \left( \int_a^b (f(x))^2 \right)^{1/2} \left( \int_a^b (g(x))^2 \right)^{1/2}$$

(obtenida al aplicar el mismo resultado al producto escalar (III) del ejemplo 8.1).

Estas desigualdades son difíciles de probar directamente, pero inmediatas gracias a la desigualdad de Schwartz. La siguiente desigualdad es clásica, y también se prueba usando la desigualdad de Schwartz.

**Proposición 8.2 (Desigualdad triangular)**

$\forall \mathbf{x}, \mathbf{y}$ de un e.e. $E$, se verifica:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

*Demostración:*

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle + \|\mathbf{y}\|^2 \\ \text{(Des. Schwartz)} &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\cdot\|\mathbf{y}\| + \|\mathbf{y}\|^2 = \left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right) \end{aligned}$$

de donde se sigue el resultado.

■

**Nota 8.3**

En la sección 3.3.2 definimos una norma como un aplicación verificando

(I) $\|\mathbf{x}\| \geq 0$, y $\|\mathbf{x}\| = 0$ si y sólo si $\mathbf{x} = \mathbf{0}$.

(II) $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|$, $\forall\lambda \in \mathbb{K}$.

(III) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Las dos primeras propiedades son consecuencia inmediata de la Definición 8.2 y las propiedades del producto escalar, mientras que la tercera es precisamente la desigualdad triangular. Es decir, la norma dada en la Definición 8.2 cumple con las propiedades anteriores.

### 8 1 1 Matriz de un producto escalar

Una vez más, el hecho de trabajar en un espacio vectorial de dimensión finita en el que existen bases nos permite acudir a las coordenadas para trabajar con los productos escalares.

En concreto, si $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ es una base de $E$ y $\mathbf{x}, \mathbf{y} \in E$ sabemos que

$$\mathbf{x} = \sum_{i=1}^n x_i\mathbf{u}_i, \quad \mathbf{y} = \sum_{i=1}^n y_i\mathbf{u}_i.$$

Entonces, por las propiedades de linealidad del producto escalar tenemos que

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y}\rangle &= \left\langle \sum_{i=1}^n x_i\mathbf{u}_i, \sum_{j=1}^n y_j\mathbf{u}_j \right\rangle = \sum_{i=1}^n x_i \left\langle \mathbf{u}_i, \sum_{j=1}^n y_j\mathbf{u}_j \right\rangle = \sum_{i=1}^n\sum_{j=1}^n x_i y_j \langle \mathbf{u}_i, \mathbf{u}_j\rangle \\ &= \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} P_{\mathcal{B}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{x}^T P_{\mathcal{B}}\mathbf{y} \end{aligned}$$

donde $(P_{\mathcal{B}})_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle$, es la denominada *matriz del producto escalar* o *matriz de Gram.*[2] Debido a la simetría del producto escalar está claro que la matriz $P_{\mathcal{B}}$ es simétrica. Nótese también que para el producto escalar habitual $P_{\mathcal{B}} = I_n$.

Dicho de otra forma, si conocemos la matriz de Gram respecto de una base, el producto escalar se convierte en una operación matricial que involucra las coordenadas de los vectores respecto de esa misma base.

**Ejemplo 8.2**

Consideremos el espacio euclídeo $\mathbb{P}^2_{\mathbb{R}}$ de polinomios de grado menor o igual con coeficientes reales dotado del producto escalar

$$\langle p(t), q(t) \rangle = \int_{-1}^{1} p(t)q(t)\, dt$$

La matriz de este producto escalar respecto de la base canónica $\mathcal{B}_c = \{1, t, t^2\}$ se calcula:

$$\langle 1, 1 \rangle = \int_{-1}^{1} 1\, dt = 2, \quad \langle 1, t \rangle = \int_{-1}^{1} t\, dt = 0, \quad \langle 1, t^2 \rangle = \int_{-1}^{1} t^2\, dt = \tfrac{2}{3},$$
$$\langle t, t \rangle = \int_{-1}^{1} t^2\, dt = \tfrac{2}{3}, \quad \langle t, t^2 \rangle = \int_{-1}^{1} t^3\, dt = 0, \quad \langle t^2, t^2 \rangle = \int_{-1}^{1} t^4\, dt = \tfrac{2}{5}$$

es decir

$$P_{\mathcal{B}_c} = \begin{pmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{pmatrix}$$

Veamos cómo usar esta matriz. Supongamos que queremos calcular el producto escalar de dos vectores de $\mathbb{P}^2_{\mathbb{R}}$, por ejemplo, $1 + 2t^2$ y $1 - 3t - t^2$. Podemos usar la definición del producto escalar:

$$\langle 1+2t^2, 1-3t-t^2 \rangle = \int_{-1}^{1} (1+2t^2)(1-3t-t^2)\, dt = \int_{-1}^{1} (1-3t+t^2-6t^3-2t^4)\, dt = \tfrac{28}{15}$$

pero también podemos usar las coordenadas de estos vectores en una determinada base y la matriz de Gram del producto escalar respecto de esa misma base. En este caso quedaría:

$$\langle 1 + 2t^2, 1 - 3t - t^2 \rangle = \begin{pmatrix} 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{pmatrix} \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix} = \tfrac{28}{15}$$

[2] Recibe su nombre del matemático danés Jørgen Pedersen Gram.

8 2

## ORTOGONALIDAD

La sección anterior nos ha mostrado qué es un producto escalar, la definición de norma y algunas propiedades interesantes; pero sin duda alguna, el concepto fundamental vinculado al producto escalar es el de *ortogonalidad*.

**Definición 8.3**

Sea $E$ un e.e.

(I) Dos vectores $\mathbf{x}$ e $\mathbf{y}$ se dicen *ortogonales* respecto de un producto escalar si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Se notará $\mathbf{x} \perp \mathbf{y}$.

(II) Un conjunto de vectores $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ se dice ortogonal (dos a dos) si $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ si $i \neq j$.

(III) Un vector $\mathbf{x}$ es ortogonal a un conjunto $S$ si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\forall \mathbf{y} \in S$.

**Ejemplo 8.3**

(I) Si consideramos en $\mathbb{R}^2$ un vector $\mathbf{x} = (\alpha, \beta)$, entonces el vector $\mathbf{y} = (-\beta, \alpha)$ es ortogonal a $\mathbf{x}$, pues

$$\mathbf{x} \cdot \mathbf{y} = (\alpha, \beta) \cdot (-\beta, \alpha) = -\alpha\beta + \beta\alpha = 0$$

(II) Si consideramos $\mathbb{R}^n$ con el producto escalar, la base canónica es un conjunto ortogonal.

(III) En $\mathcal{C}([-\pi, \pi])$ con el producto escalar

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x)\, dx$$

el conjunto $\{\operatorname{sen}(mx), \cos(nx)\}_{m,n \in \mathbb{N}}$ es ortogonal (dos a dos).

Cuando tratamos con vectores de $\mathbb{R}^2$ o $\mathbb{R}^3$ y el producto escalar habitual, la ortogonalidad coincide con el concepto de perpendicularidad que el lector probablemente conocerá. Sin embargo, el concepto de ortogonalidad es mucho más amplio y puede establecerse en cualquier espacio euclídeo, por ejemplo,

en un espacio de funciones. El concepto de funciones ortogonales (o incluso el de polinomios ortogonales) tiene importantes e interesantes aplicaciones que se escapan fuera del alcance de este texto.

Una consecuencia inmediata de la ortogonalidad es la independencia lineal:

**Proposición 8.3**

Si $\mathbf{x}_1, \dots, \mathbf{x}_n$ son ortogonales entre sí y no nulos entonces son l.i.

*Demostración:*
Veamos que la combinación lineal nula de estos vectores,

$$\alpha_1\mathbf{x}_1 + \dots + \alpha_n\mathbf{x}_n = \mathbf{0}$$

implica que los escalares deben ser nulos. Haciendo el producto escalar para cada uno de los vectores $\mathbf{x}_j$ se tiene

$$\langle \alpha_1\mathbf{x}_1 + \dots + \alpha_n\mathbf{x}_n, \mathbf{x}_j\rangle = \langle \mathbf{0}, \mathbf{x}_j\rangle = 0$$

Y usando la linealidad,

$$\alpha_1\langle \mathbf{x}_1, \mathbf{x}_j\rangle + \dots + \alpha_n\langle \mathbf{x}_n, \mathbf{x}_j\rangle = 0 \Rightarrow \alpha_j\langle \mathbf{x}_j, \mathbf{x}_j\rangle = 0 \Rightarrow \alpha_j = 0, \ \forall j$$

Nótese que es fundamental que ninguno de los vectores sea nulo.

■

Otra consecuencia de la ortogonalidad es la siguiente versión del Teorema de Pitágoras:

**Proposición 8.4 (Teorema de Pitágoras)**

Si $\mathbf{x} \perp \mathbf{y}$ entonces

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

La demostración se propone como ejercicio al lector. En la figura 8.1 representamos el resultado en el contexto de vectores en un plano. La suma de dos vectores ortogonales corresponde a un vector que coincide con la hipotenusa del triángulo rectángulo que forman, mientras que las normas corresponden a las longitudes de los mismos; es decir, en un triángulo rectángulo, la suma de los cuadrados de las longitudes de los lados ortogonales es igual al cuadrado de la longitud de la hipotenusa, esto es, el enunciado clásico del Teorema de Pitágoras.

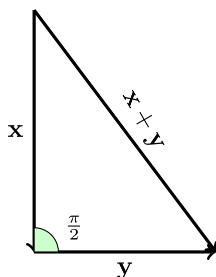



Figura 8.1: Ilustración del Teorema de Pitágoras

### 8.2.1 Bases ortonormales

Un concepto habitual en los espacios euclídeos es el de *base ortonormal*.

**Definición 8.4**

Se dice que una base $\mathcal{B}$ es *ortogonal* si los elementos que la forman son ortogonales dos a dos. Si además los vectores son de norma uno respecto del mismo producto escalar, se dirá que $\mathcal{B}$ es *ortonormal*.

**Teorema 8.1**

En todo espacio euclídeo de dimensión finita existen bases ortonormales.

La demostración de este resultado es consecuencia inmediata del siguiente teorema.

**Teorema 8.2 (Gram-Schmidt, método de ortogonalización[3])**

Sea $\mathbf{x}_1, \dots, \mathbf{x}_n, \dots$, una sucesión (finita o infinita) de vectores linealmente independientes. Denotemos por $L_k = L(\mathbf{x}_1, \dots, \mathbf{x}_k)$ el espacio generado por los $k$ primeros vectores. Entonces existen $\mathbf{y}_1, \dots, \mathbf{y}_n, \dots$ tales que

(I) $L_k = L(\mathbf{y}_1, \dots, \mathbf{y}_k)$.

(II) $\mathbf{y}_{k+1} \perp L_k$, $\forall k \geq 1$.

Antes de pasar a la demostración conviene entender bien lo que significa este resultado. Dado un conjunto de vectores linealmente independientes, es posible encontrar un nuevo conjunto de vectores que "generan lo mismo" que los vectores originales (propiedad (i)), y que además son ortogonales entre sí, dos a dos (consecuencia de la propiedad (ii)). Es decir, dado un conjunto de vectores, podemos ortogonalizar dicho conjunto, con vectores que son "esencialmente iguales", en el sentido de que generan los mismos subespacios.

*Demostración:*

La demostración de este resultado es constructiva, es decir, nos dice cómo, a partir de un cierto conjunto de vectores, podemos construir un nuevo conjunto que genera lo mismo, y que son ortogonales. Dicho método es conocido como el *método de ortogonalización de Gram-Schmidt.*

Para construir estos vectores, definimos inductivamente la sucesión de vectores $\mathbf{y}_j$ del siguiente modo:

$$\mathbf{y}_1 = \mathbf{x}_1, \quad \mathbf{y}_{k+1} = \mathbf{x}_{k+1} - \sum_{j=1}^{k} \frac{\langle \mathbf{y}_j, \mathbf{x}_{k+1} \rangle}{\|\mathbf{y}_j\|^2} \mathbf{y}_j, \text{ para } k \geq 1 \tag{8.1}$$

Veamos en primer lugar que esta sucesión satisface (i), usando el principio de inducción:

- Para $k = 1$ es evidente que $L_1 = L(\mathbf{y}_1)$.
- Supongamos ahora que $L_k = L(\mathbf{y}_1, \dots, \mathbf{y}_k)$ y veamos que también se tiene el resultado para $k + 1$. Obsérvese que de (8.1), $\mathbf{y}_{k+1} = \mathbf{x}_{k+1} - \mathbf{v}$, con $\mathbf{v} \in L_k$ (por hipótesis de inducción), luego $\mathbf{y}_{k+1} \in L_{k+1}$. De aquí es evidente que $L_{k+1} = L(\mathbf{y}_1, \dots, \mathbf{y}_{k+1})$.

Probemos ahora (ii), también por inducción.

- $\mathbf{y}_2 \perp L_1$ pues

$$\langle \mathbf{y}_2, \mathbf{y}_1 \rangle = \langle \mathbf{x}_2, \mathbf{x}_1 \rangle - \frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{\|\mathbf{x}_1\|^2} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle = 0$$

- Supongamos ahora que $\mathbf{y}_k \perp L_{k-1}$ y probemos que $\mathbf{y}_{k+1}$ es ortogonal a $L_k$. Para ello debemos ver que $\langle \mathbf{y}_{k+1}, \mathbf{y}_j \rangle = 0$, $1 \leq j \leq k$. Entonces,

$$\begin{aligned} \langle \mathbf{y}_{k+1}, \mathbf{y}_j \rangle &= \left\langle \mathbf{x}_{k+1} - \sum_{i=1}^{k} \frac{\langle \mathbf{y}_i, \mathbf{x}_{k+1} \rangle}{\|\mathbf{y}_i\|^2} \mathbf{y}_i, \mathbf{y}_j \right\rangle \\ &= \langle \mathbf{x}_{k+1}, \mathbf{y}_j \rangle - \sum_{i=1}^{k} \frac{\langle \mathbf{y}_i, \mathbf{x}_{k+1} \rangle}{\|\mathbf{y}_i\|^2} \langle \mathbf{y}_i, \mathbf{y}_j \rangle \end{aligned}$$

[3] El nombre se debe a Jørgen Gram y al ruso (nacido en la actual Estonia) Erhard Schmidt, aunque el resultado, ya usado por Cauchy en 1836, parece ser debido al francés Pierre-Simon Laplace.

Por hipótesis de inducción sabemos que $\langle \mathbf{y}_i, \mathbf{y}_j \rangle = 0$ si $i \neq j$, con $1 \leq i, j \leq k$; luego

$$\langle \mathbf{y}_{k+1}, \mathbf{y}_j \rangle = \langle \mathbf{x}_{k+1}, \mathbf{y}_j \rangle - \langle \mathbf{y}_j, \mathbf{x}_{k+1} \rangle = 0$$

■

Veamos algunos ejemplos.

**Ejemplo 8.4**

(I) Consideremos el conjunto $\{(0,1,2),(1,0,0),(0,1,1)\}$. Ortonormalizar respecto del producto escalar habitual en $\mathbb{R}^3$.

Aplicamos el método de Gram-Schmidt del siguiente modo,

$$\begin{aligned}
\mathbf{y}_1 &= (0,1,2) \\
\mathbf{y}_2 &= \mathbf{x}_2 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_2}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 = (1,0,0) - \frac{(1,0,0)\cdot(0,1,2)}{\|(0,1,2)\|^2}(0,1,2) = (1,0,0) \\
\mathbf{y}_3 &= \mathbf{x}_3 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_3}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 - \frac{\mathbf{y}_2 \cdot \mathbf{x}_3}{\|\mathbf{y}_2\|^2}\mathbf{y}_2 \\
&= (0,1,1) - \frac{(0,1,2)\cdot(0,1,1)}{\|(0,1,2)\|^2}(0,1,2) - \frac{(1,0,0)\cdot(0,1,1)}{\|(1,0,0)\|^2}(1,0,0) \\
&= (0, \tfrac{2}{5}, -\tfrac{1}{5})
\end{aligned}$$

Así pues, el conjunto $\{(0,1,2),(1,0,0),(0,\frac{2}{5},-\frac{1}{5})\}$ es ortogonal. Nótese que puesto que $(0,1,2)$ y $(1,0,0)$ ya son ortogonales, el método no los modifica. Por otro lado, es evidente que los subespacios

$$L_1 = L((0,1,2)), \quad L_2 = L((0,1,2),(1,0,0)),$$
$$L_3 = L((0,1,2),(1,0,0),(0,\tfrac{2}{5},-\tfrac{1}{5}))$$

son los mismos que

$$L_1 = L((0,1,2)), \quad L_2 = L((0,1,2),(1,0,0)),$$
$$L_3 = L((0,1,2),(1,0,0),(0,1,1))$$

Puesto que además, los vectores iniciales forman base de $\mathbb{R}^3$ (son tres vectores independientes), podemos obtener una base ortonormal dividiendo cada uno de ellos por su norma:

$$\|(0,1,2)\| = \sqrt{(0,1,2)\cdot(0,1,2)} = \sqrt{1+4} = \sqrt{5}, \quad \|(1,0,0)\| = 1,$$
$$\|(0,\tfrac{2}{5},-\tfrac{1}{5})\| = \sqrt{\tfrac{1}{5}}$$

Es decir, el conjunto $\{(0,\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}),(1,0,0),(0,\frac{2\sqrt{5}}{5},-\frac{\sqrt{5}}{5})\}$ es una base ortonormal de $\mathbb{R}^3$.

(II) Ortonormalizar los vectores $\{1,t,t^2,t^3\}$ de $\mathbb{P}_{\mathbb{R}}[t]$ respecto del producto escalar

$$\langle p,q\rangle = \int_{-1}^{1} p(t)q(t)\,dt$$

En el ejemplo anterior hemos acudido directamente a la expresión (8.1) que nos proporciona el método de Gram-Schmidt. Sin embargo, en la práctica no es necesario conocer esta fórmula "de memoria", sino observar cómo se construye cada nuevo elemento de la sucesión para ortogonalizar un conjunto cualquiera. Para el caso que nos ocupa, está claro que el primer elemento de la nueva sucesión es el primer elemento que ya tenemos:

$$\mathbf{y}_1 = 1$$

Ahora, para calcular el siguiente elemento basta ver que se trata del elemento de la sucesión dada menos una combinación lineal de todos los elementos que acabamos de construir. Así,

$$\mathbf{y}_2 = \mathbf{x}_2 - \alpha\mathbf{y}_1$$

con la condición de que $\langle \mathbf{y}_2,\mathbf{y}_1\rangle = 0$. Es decir, $\mathbf{y}_2 = t-\alpha$, siendo $\alpha$ tal que

$$\langle t-\alpha,1\rangle = \int_{-1}^{1}(t-\alpha)\,dt = 0 \Rightarrow \alpha = 0$$

luego $\mathbf{y}_2 = t$. Nótese que puesto que $\mathbf{x}_1 = 1$ y $\mathbf{x}_2 = t$ son ya ortogonales, el método Gram-Schmidt no modifica dichos vectores.

Del mismo modo, $\mathbf{y}_3$ se construye como $\mathbf{y}_3 = \mathbf{x}_3 - \alpha_1\mathbf{y}_1 - \alpha_2\mathbf{y}_2$, con las condiciones adicionales $\langle \mathbf{y}_3,\mathbf{y}_1\rangle = \langle \mathbf{y}_3,\mathbf{y}_2\rangle = 0$. Imponiendo ambas condiciones:

$$\begin{aligned}\langle t^2-\alpha_1-\alpha_2 t,1\rangle &= \int_{-1}^{1}(t^2-\alpha_1-\alpha_2 t)\,dt = 0\\ \langle t^2-\alpha_1-\alpha_2 t,t\rangle &= \int_{-1}^{1}(t^2-\alpha_1-\alpha_2 t)t\,dt = 0\end{aligned} \Rightarrow \left\{\begin{array}{l}\frac{2}{3}-2\alpha_1 = 0\\ -\frac{2}{3}\alpha_2 = 0\end{array}\right.$$

luego $\alpha_1 = \frac{1}{3}$ y $\alpha_2 = 0$. Es decir, $\mathbf{y}_3 = t^2 - \frac{1}{3}$.

Finalmente, $\mathbf{y}_4 = t^3-\beta_1-\beta_2 t-\beta_3(t^2-\frac{1}{3})$, con las condiciones $\langle \mathbf{y}_4,\mathbf{y}_j\rangle = 0$,

$1 \leq j \leq 3$, es decir

$$\langle t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3}), 1\rangle = \int_{-1}^{1} \left(t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3})\right)\, dt = 0$$

$$\langle t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3}), t\rangle = \int_{-1}^{1} \left(t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3})\right) t\, dt = 0$$

$$\langle t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3}), t^2 - \tfrac{1}{3}\rangle$$

$$= \int_{-1}^{1} \left(t^3 - \beta_1 - \beta_2 t - \beta_3(t^2 - \tfrac{1}{3})\right)\left(t^2 - \tfrac{1}{3}\right)\, dt = 0$$

de donde se obtiene el sistema

$$\begin{cases} -2\beta_1 = 0 \\ \frac{2}{5} - \frac{2}{3}\beta_2 = 0 \\ -\frac{8}{45}\beta_3 = 0 \end{cases}$$

cuyas soluciones son $\beta_1 = \beta_3 = 0$, $\beta_2 = \frac{3}{5}$. Así pues, $\mathbf{y}_4 = t^3 - \frac{3}{5}t$. Es importante observar que el cálculo de las integrales anteriores se puede simplificar bastante si observamos que en diversos momentos estamos calculando integrales del tipo

$$\int_{-1}^{1} (t^2 - \tfrac{1}{3})\, dt = \langle 1, t^2 - \tfrac{1}{3}\rangle = 0 \quad \text{ó} \quad \int_{-1}^{1} t(t^2 - \tfrac{1}{3})\, dt = \langle t, t^2 - \tfrac{1}{3}\rangle = 0$$

que son nulas debido a la ortogonalidad entre estos polinomios.

Finalmente, el conjunto $\{1, t, t^2 - \frac{1}{3}, t^3 - \frac{3}{5}t\}$ es ortogonal, y calculando sus normas:

$$\|1\| = \left(\int_{-1}^{1} 1\, dt)\right)^{\frac{1}{2}} = \sqrt{2}, \ \|t\| = \left(\int_{-1}^{1} t^2\, dt)\right)^{\frac{1}{2}} = \sqrt{\tfrac{2}{3}}$$

$$\|t^2 - \tfrac{1}{3}\| = \sqrt{\tfrac{8}{45}}, \ \|t^3 - \tfrac{3}{5}t\| = \sqrt{\tfrac{8}{175}}$$

obtenemos el siguiente conjunto ortonormal:

$$\left\{\tfrac{1}{\sqrt{2}}, \sqrt{\tfrac{3}{2}}t, \sqrt{\tfrac{45}{8}}(t^2 - \tfrac{1}{3}), \sqrt{\tfrac{175}{8}}(t^3 - \tfrac{3}{5}t)\right\}$$

En particular, obsérvese que con el producto escalar dado, la base canónica de $\mathbb{P}^3_{\mathbb{R}}$ no es ortonormal.

**Nota 8.4**

(I) Obsérvese que en una base ortonormal $\mathcal{B}$, la matriz del producto escalar $P_{\mathcal{B}} = I_n$. Recíprocamente, si la matriz de un producto escalar es la identidad, entonces la base es ortonormal.

(II) Dada una base ortonormal $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, las coordenadas de un vector $\mathbf{x}$ respecto de esta base son

$$\mathbf{x}_{\mathcal{B}} = (\langle \mathbf{x}, \mathbf{u}_1 \rangle, \dots, \langle \mathbf{x}, \mathbf{u}_n \rangle)$$

pues si $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{u}_i$, multiplicando escalarmente por cada vector $\mathbf{u}_j$ se tiene

$$\langle \mathbf{x}, \mathbf{u}_j \rangle = \left\langle \sum_{i=1}^{n} x_i \mathbf{u}_i, \mathbf{u}_j \right\rangle = \sum_{i=1}^{n} x_i \langle \mathbf{u}_i, \mathbf{u}_j \rangle = x_j$$

(III) Si $\mathcal{B}$ es una base de un e.e., $P_{\mathcal{B}}$ la matriz del producto escalar y $\mathcal{B}' = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ es una base ortonormal respecto del producto escalar, entonces la matriz $A$ cuyas columnas corresponden a las coordenadas de los vectores $\mathbf{u}_i$ respecto de la base $\mathcal{B}$ verifica que $A^T P_{\mathcal{B}} A = I_n$.

En particular, si consideramos el producto escalar habitual en $\mathbb{R}^n$ y la base canónica, cualquier matriz $A$ cuyas columnas correspondan a las coordenadas (canónicas) de los vectores de una base ortonormal verifica que $A^T A = I_n$. Esto motiva la siguiente definición.

**Definición 8.5**

Se dice que una matriz $A \in \mathcal{M}_n(\mathbb{K})$ es *ortogonal* si $A^T A = I_n$, o equivalentemente, $A^T = A^{-1}$

## 8.2.2 Subespacios ortogonales

El concepto de ortogonalidad puede ser extendido también a subespacios:

**Definición 8.6**

Dos subespacios $W_1$, $W_2$ se dicen *ortogonales*, y se notará $W_1 \perp W_2$, si todos los vectores de $W_1$ son ortogonales a todos los vectores de $W_2$.

Como es habitual en los espacios vectoriales, lo que sucede con una base es extensible a todo el subespacio.

**Proposición 8.5**

$W_1 \perp W_2$ si y sólo si los vectores de una base de $W_1$ son ortogonales a los vectores de una base de $W_2$.

*Demostración:*
La implicación hacia la derecha es trivial. Veamos la implicación contraria. Sea $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ una base de $W_1$ y $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ una base de $W_2$, de forma que los elementos de la base de $W_1$ son ortogonales a los elementos de la base de $W_2$. Consideremos ahora $\mathbf{x} \in W_1$ e $\mathbf{y} \in W_2$. Por la condición de base,

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{u}_i, \quad \mathbf{y} = \sum_{j=1}^{m} y_j \mathbf{v}_j$$

Entonces, de la linealidad del producto escalar

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0$$

pues $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0 \ \forall i, j$, de modo que $\mathbf{x} \perp \mathbf{y}$. De la arbitrariedad de los vectores se obtiene el resultado.

■

**Proposición 8.6**

$W_1 \perp W_2 \Rightarrow W_1 \cap W_2 = \{\mathbf{0}\}$.

*Demostración:*
Está claro que si $\mathbf{x} \in W_1 \cap W_2$, dado que $W_1 \perp W_2$ se debe tener que $\langle \mathbf{x}.\mathbf{x} \rangle = \mathbf{0}$, de lo que se deduce que $\mathbf{x} = \mathbf{0}$.

■

**Definición 8.7**

Sea $W$ un subespacio vectorial de un e.e. $E$. Se define el *ortogonal de* $W$, también denominado *complemento ortogonal* de $W$, y se notará por $W^{\perp}$, como

$$W^{\perp} = \{\mathbf{y} \in E : \mathbf{x} \perp \mathbf{y},\ \forall \mathbf{x} \in W\}$$

Es decir, el complemento ortogonal de un conjunto $W$ está formado por los vectores que son ortogonales (simultáneamente) a todos los vectores de $W$.

**Proposición 8.7**

$W^{\perp}$ es un subespacio vectorial.

La demostración se propone como ejercicio al lector.

**Proposición 8.8**

$(W^{\perp})^{\perp} = W$.

*Demostración:*
Sea $\mathbf{x} \in W$. Entonces si $\mathbf{y} \in W^{\perp}$ entonces $\mathbf{x} \perp \mathbf{y}$, es decir, $\mathbf{x} \in (W^{\perp})^{\perp}$. De aquí se concluye que $W \subset (W^{\perp})^{\perp}$.

Recíprocamente, si $\mathbf{x} \in (W^{\perp})^{\perp}$ entonces $\mathbf{x} \perp \mathbf{y}$ para todo $\mathbf{y} \in W^{\perp}$. Esto es, $\mathbf{x} \in W$, lo que implica que $(W^{\perp})^{\perp} \subset W$. De la doble inclusión se obtiene el resultado deseado.

■

### Cálculo del ortogonal a un subespacio

La Proposición 8.5 nos indica que para obtener la ortogonalidad a un subespacio, basta trabajar con una base del subespacio. Esta es la idea central para obtener el ortogonal de un subespacio.

De forma general podemos afirmar que si $W$ tiene una base formada por los vectores $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, entonces $W^{\perp}$ viene dado por el sistema de ecuaciones implícitas

$$\langle \mathbf{x}, \mathbf{u}_1 \rangle = 0,\ \dots,\ \langle \mathbf{x}, \mathbf{u}_n \rangle = 0$$

Es más, si usamos el producto escalar habitual de $\mathbb{R}^n$, y $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ es una base de $W$, entonces, la matriz cuyas filas están formadas por las coordenadas

de los vectores de la base corresponde a la matriz de un sistema de ecuaciones implícitas que define a $W^\perp$.

**Ejemplo 8.5**

(I) Para encontrar el ortogonal de $W = L((1,1,0),(0,0,1))$ respecto al producto escalar habitual de $\mathbb{R}^3$ bastará imponer ortogonalidad respecto a la base de $W$:

$$\mathbf{x} \perp (1,1,0), \quad \mathbf{x} \perp (0,0,1)$$

Si ponemos $\mathbf{x} = (x_1, x_2, x_3)$ entonces el requerimiento anterior nos lleva al sistema

$$\left.\begin{array}{rcl} x_1 + x_2 & = & 0 \\ x_3 & = & 0 \end{array}\right\}$$

que resulta un sistema de ecuaciones implícitas de $W^\perp$. Resolviendo dicho sistema encontramos una base de $W^\perp$; en este caso $W^\perp = L((1,-1,0))$. Nótese que la matriz del sistema de ecuaciones que define a $W^\perp$ es

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

cuyas filas corresponden a las coordenadas de los vectores de la base de $W$.

(II) Encontrar el ortogonal de $W = \{\mathbf{x} \in \mathbb{R}^3 : x_1 + 2x_2 - x_3 = 0\}$.

En una primera aproximación, para calcular $W^\perp$ procederíamos como en (I), es decir, calcularíamos una base de $W$ (resolviendo el sistema) y posteriormente la emplearíamos para construir un sistema de ecuaciones implícitas de $W^\perp$. Sin embargo es más sencillo prestar atención al hecho de que $\mathbf{x} \in W$ si satisface la ecuación

$$x_1 + 2x_2 - x_3 = 0 \Leftrightarrow (x_1, x_2, x_3) \cdot (1, 2, -1) = 0$$

Es decir, todos los vectores de $W$ deben ser ortogonales a $(1,2,-1)$, de manera que este vector forma una base de $W^\perp$.

Atendiendo a este último ejemplo, si estamos usando el producto escalar habitual en $\mathbb{R}^n$ y tenemos un subespacio $W$ generado por las soluciones del sistema homogéneo $A\mathbf{x} = \mathbf{0}$, entonces una base de $W^\perp$ estará formada por los vectores cuyas coordenadas corresponden a las filas de la matriz $A$.

Si estamos usando un producto escalar distinto, cuya matriz respecto de la base con la que trabajemos es $P_{\mathcal{B}}$, entonces la matriz $AP_{\mathcal{B}}$, donde $A$ es la matriz cuyas filas están formadas por las coordenadas respecto de la base $\mathcal{B}$ de los vectores de la base de $W$, es la matriz de un sistema de ecuaciones implícitas que define a $W^{\perp}$.

**Ejemplo 8.6**

Calculemos ahora el subespacio ortogonal a $W = L(\{2+t, 1-t^2\})$ en $\mathbb{P}^2_{\mathbb{R}}$. Dado que la base de $W$ está formada por los polinomios $2+t$ y $1-t^2$, una primera opción para obtener $W^{\perp}$ sería obtener los polinomios $x_1 + x_2 t + x_3 t^2$ tales que

$$\langle x_1 + x_2 t + x_3 t^2, 2+t\rangle = 0, \quad \langle x_1 + x_2 t + x_3 t^2, 1-t^2\rangle = 0$$

es decir,

$$\left\{ \begin{array}{l} \displaystyle\int_{-1}^{1} (1 + x_2 t + x_3 t^2)(2+t)\, dt = 4x_1 + \tfrac{2}{3}x_2 + \tfrac{4}{3}x_3 = 0 \\ \displaystyle\int_{-1}^{1} (x_1 + x_2 t + x_3 t^2)(1-t^2)\, dt = \tfrac{4}{3}x_1 + \tfrac{4}{15}x_3 = 0 \end{array} \right.$$

cuya solución es $(1, 4, -5)$, que corresponde al polinomio $1 + 4t - 5t^2$.

En lugar de calcular los productos escalares mediante integrales, podemos hacer uso de la matriz de Gram de este producto escalar respecto de la base canónica, ya obtenida en el ejemplo 8.2. En tal caso, $W^{\perp}$ viene dado por un sistema de ecuaciones homogéneo cuya matriz es

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{pmatrix} = \begin{pmatrix} 4 & \frac{2}{3} & \frac{4}{3} \\ \frac{4}{3} & 0 & \frac{4}{15} \end{pmatrix}$$

que obviamente coincide con el obtenido anteriormente.

### 8 2 3 Proyección ortogonal

Una vez introducido el concepto de complemento ortogonal de un subespacio, vamos a probar un resultado que nos permitirá definir la *proyección ortogonal*.

**Proposición 8.9**

Si $W$ es un subespacio vectorial de un e.e. $E$, entonces $W \oplus W^{\perp} = E$.

*Demostración:*
Por la Proposición 8.6 ya sabemos que $W \cap W^{\perp} = \{\mathbf{0}\}$. Veamos ahora que $W + W^{\perp} = E$.

Consideramos una base de $W$, $\{\mathbf{u}_1, \dots \mathbf{u}_n\}$ y la completamos hasta obtener una base de $E$ que posteriormente ortogonalizamos mediante el método de Gram-Schmidt. De este modo tenemos una base ortogonal formada por unos vectores $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{v}_{n+1}, \dots, \mathbf{v}_m\}$ tales que $L(\mathbf{u}_1, \dots, \mathbf{u}_n) = L(\mathbf{v}_1, \dots, \mathbf{v}_n) = W$.

Está claro que si $\mathbf{z} \in E$, entonces

$$\mathbf{z} = \sum_{i=1}^{m} z_i \mathbf{v}_i = \sum_{i=1}^{n} z_i \mathbf{v}_i + \sum_{j=n+1}^{m} z_j \mathbf{v}_j = \mathbf{x} + \mathbf{y}$$

Obviamente $\mathbf{x} \in W$. Por otro lado, $\mathbf{y} \perp \mathbf{v}_i$, $1 \leq i \leq n$, pues

$$\langle \mathbf{y}, \mathbf{v_i} \rangle = \left\langle \sum_{j=n+1}^{m} z_j \mathbf{v}_j, \mathbf{v}_i \right\rangle = \sum_{j=n+1}^{m} z_j \langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0, \quad i = 1, \dots n$$

de modo que $\mathbf{y} \in W^{\perp}$ (pues es ortogonal a una base de $W$). Es decir, hemos escrito un vector cualquiera de $E$ como suma de uno de $W$ más otro de $W^{\perp}$, lo cual significa que $E = W + W^{\perp}$. Esto concluye la prueba.

■

Obsérvese que como consecuencia de este resultado todo vector $\mathbf{x} \in E$ puede escribirse **de forma única** (recuérdese el Teorema 4.13) como

$$\mathbf{x} = \mathbf{y} + \mathbf{z}, \quad \text{con } \mathbf{y} \in W \text{ y } \mathbf{z} \in W^{\perp}$$

**Definición 8.8**

Con las notaciones anteriores se dice que $\mathbf{y}$ es la *proyección ortogonal* de $\mathbf{x}$ sobre $W$, que denotaremos por $\mathcal{P}_W(\mathbf{x}) = \mathbf{y}$. Asimismo, se dice que $\mathbf{z}$ es la proyección ortogonal sobre $W^{\perp}$, que denotaremos por $\mathcal{Q}_W(\mathbf{x}) = \mathbf{z}$.

La figura 8.2 ilustra la típica proyección de un vector sobre un plano vectorial $W$. Descomponemos el vector $\mathbf{x}$ en suma de un vector de $W$, el vector $\mathbf{y}$, más otro vector del ortogonal de $W$ (que en este caso es la recta perpendicular a $W$), $\mathbf{z}$, de tal forma que $\mathbf{x} = \mathbf{y} + \mathbf{z}$. La proyección ortogonal sobre $W$ es en este caso el vector $\mathbf{y}$.

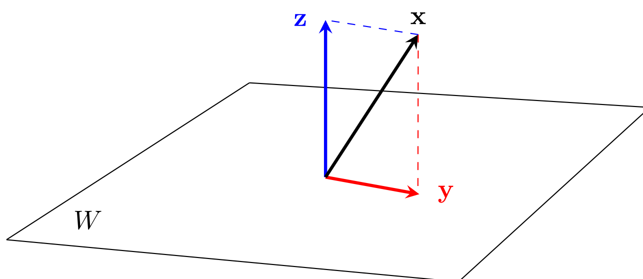



Figura 8.2: Proyección sobre un plano

### Cálculo de la proyección ortogonal a un subespacio

Procedamos ahora a calcular la proyección ortogonal de un vector $\mathbf{x}$ sobre un subespacio $W$.

Dada una base de $W$, $\{\mathbf{e}_1, \dots, \mathbf{e}_r\}$, sabemos que la proyección ortogonal de un vector $\mathbf{x} \in E$ es un vector $\mathbf{y} \in W$, de manera que $\mathbf{y} = \sum_{i=1}^{r} \lambda_i \mathbf{e}_i$, donde $\lambda_i$ son escalares que vamos a determinar.

Por otra parte, como $\mathbf{x} = \mathbf{y} + \mathbf{z}$ con $\mathbf{z} \in W^\perp$, eso significa que $\mathbf{x} - \mathbf{y} \in W^\perp$. Es decir, el vector $\mathbf{z} = \mathbf{x} - \mathbf{y}$, conocido como el *error*,[4] es ortogonal a $W$. Por tanto debe ser ortogonal a su base, es decir

$$\left\langle \mathbf{x} - \sum_{i=1}^{r} \lambda_i \mathbf{e}_i, \mathbf{e}_j \right\rangle = 0, \quad j = 1 \dots r \tag{8.2}$$

De este modo, (8.2) es un sistema de $r$ ecuaciones con $r$ incógnitas $(\lambda_1, \dots \lambda_r)$ a partir del cual obtenemos los escalares que nos determinan la proyección.

**Ejemplo 8.7**

(I) Determinar la proyección del vector $\mathbf{x} = (0, -2, 1)$ sobre $W = L((1, 1, 0))$.

Denotando por $\mathcal{P}_W(\mathbf{x})$ al vector buscado, está claro que $\mathcal{P}_W(\mathbf{x})$ debe pertenecer al subespacio $W$, de modo que $\mathcal{P}_W(\mathbf{x}) = \lambda(1, 1, 0)$. Por otra parte, el error $(0, -2, 1) - \lambda(1, 1, 0)$ debe ser ortogonal a $W$, de manera que

$$((0, -2, 1) - \lambda(1, 1, 0)) \cdot (1, 1, 0) = 0$$
$$\Rightarrow (0, -2, 1) \cdot (1, 1, 0) - \lambda(1, 1, 0) \cdot (1, 1, 0) = 0$$

[4] Se denomina así pues es la diferencia entre el vector y su proyección.

de lo que se deduce que $\lambda = -1$ y la proyección buscada será $\mathcal{P}_W(\mathbf{x}) = (-1, -1, 0)$.

(II) Determinar la proyección ortogonal del vector $t^2$ respecto del subespacio $W = L(\{1, t\})$ en $\mathbb{P}_\mathbb{R}$, dotado del producto escalar

$$\langle p, q\rangle = \int_{-1}^{1} p(t)q(t)\, dt$$

Al igual que antes, la proyección $\mathcal{P}_W(t^2)$ debe ser un vector de $W$ es decir, será de la forma $\lambda_1 \cdot 1 + \lambda_2 t$. El vector error, esto es

$$\mathcal{Q}_W(t^2) = t^2 - \lambda_1 - \lambda_2 t$$

estará en $W^\perp$, por tanto será ortogonal a una base de $W$. De este modo,

$$\begin{array}{l} \langle t^2 - \lambda_1 - \lambda_2 t, 1\rangle = 0 \\ \langle t^2 - \lambda_1 - \lambda_2 t, t\rangle = 0 \end{array} \Rightarrow \left\{ \begin{array}{l} \displaystyle\int_{-1}^{1} (t^2 - \lambda_1 - \lambda_2 t)\, dt = 0 \\ \displaystyle\int_{-1}^{1} (t^3 - \lambda_1 t - \lambda_2 t^2)\, dt = 0 \end{array} \right.$$

de donde se obtiene el sistema

$$\left. \begin{array}{rcl} \frac{2}{3} - 2\lambda_1 & = & 0 \\ -\frac{2}{3}\lambda_2 & = & 0 \end{array} \right\} \Rightarrow \lambda_1 = \frac{1}{3},\ \ \lambda_2 = 0$$

Así pues la proyección ortogonal es $\mathcal{P}_W(t^2) = \frac{1}{3}$.

(III) Calculemos la proyección del vector $\mathbf{x} = (0, 1, 1)$ sobre el subespacio

$$W = L((0, 1, 0), (0, 0, 1))$$

Siguiendo los ejemplos anteriores buscamos un vector $\mathcal{P}_W(\mathbf{x}) \in W$, es decir, $\mathcal{P}_W(\mathbf{x}) = \lambda_1(0, 1, 0) + \lambda_2(0, 0, 1)$ que además verifica

$$\mathbf{x} - \mathcal{P}(\mathbf{x}) \perp W \Rightarrow \left\{ \begin{array}{l} ((0, 1, 1) - \lambda_1(0, 1, 0) - \lambda_2(0, 0, 1)) \cdot (0, 1, 0) = 0 \\ ((0, 1, 1) - \lambda_1(0, 1, 0) - \lambda_2(0, 0, 1)) \cdot (0, 0, 1) = 0 \end{array} \right.$$

de donde obtenemos el sistema $1 - \lambda_1 = 0$, $1 - \lambda_2 = 0$. luego el vector buscado es $\mathcal{P}_W(\mathbf{x}) = (0, 1, 1)$.

**Nota 8.5**

De (III) del ejemplo anterior podemos extraer dos consecuencias importantes.

(I) La proyección de un vector $\mathbf{x}$ sobre un subespacio $W$ cuando $\mathbf{x} \in W$ es el propio vector $\mathbf{x}$. Esto es evidente pues la única forma de escribir $\mathbf{x}$ como suma de un vector de $W$ más otro de $W^{\perp}$ es $\mathbf{x} = \mathbf{x} + \mathbf{0}$.

(II) Si $\{\mathbf{e}_1, \dots \mathbf{e}_r\}$ es una **base ortogonal** de $W$, la proyección ortogonal es más fácil de calcular, puesto que el sistema dado en (8.2) se puede escribir

$$\langle \mathbf{x}, \mathbf{e}_j \rangle - \sum_{i=1}^{r} \lambda_i \langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0 \quad j = 1, \dots, r.$$

Teniendo en cuenta que $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ si $i \neq j$, entonces

$$\lambda_j = \frac{\langle \mathbf{x}, \mathbf{e}_j \rangle}{\|\mathbf{e}_j\|^2}, \quad j = 1, \dots, r.$$

A los $\lambda_j$ se les denomina *componentes de la proyección en la dirección* $\mathbf{e}_j$, pues la proyección resulta

$$\mathcal{P}_W(\mathbf{x}) = \sum_{j=1}^{r} \frac{\langle \mathbf{x}, \mathbf{e}_j \rangle}{\|\mathbf{e}_j\|^2} \mathbf{e}_j$$

### Matriz de proyección

Una forma alternativa de calcular la proyección de un vector $\mathbf{x}$ sobre un subespacio $W$ generado por unos vectores $\mathbf{e}_1, \dots, \mathbf{e}_k$ (que supondremos independientes) consiste en lo siguiente: siempre podemos considerar que el subespacio $W$ coincide con el subespacio imagen de una cierta aplicación de matriz $A$. Bastará para ello considerar la matriz $A$ como aquella cuyas columnas corresponden a las coordenadas de los vectores que conforman una base de $W$. Dicho de otro modo, $W = \operatorname{Im}(A)$.

Nuevamente, buscar la proyección supondrá encontrar un vector de $W$, es decir, en este caso, un vector de $\operatorname{Im}(A)$. Así pues, $\mathcal{P}_W(\mathbf{x}) = A\boldsymbol{\lambda}$, para algún vector $\boldsymbol{\lambda}$. Atendiendo al producto de matrices, podemos ver las componentes del vector $\boldsymbol{\lambda}$ como los escalares de la combinación lineal que define a la proyección.

Por otra parte sabemos que $(\mathbf{x} - A\boldsymbol{\lambda}) \perp W$, es decir, que el producto escalar de $\mathbf{x} - A\boldsymbol{\lambda}$ con cualquier vector de la base de $W$ es cero. Si estamos usando el producto escalar habitual en $\mathbb{R}^n$ es fácil darse cuenta que podemos representar

todos estos productos escalares simultáneamente escribiendo

$$A^T(\mathbf{x} - A\boldsymbol{\lambda}) = \mathbf{0} \Rightarrow A^T A\boldsymbol{\lambda} = A^T\mathbf{x} \tag{8.3}$$

Si los vectores que generan $W$, esto es, las columnas de $A$, son independientes, se tiene que la matriz $A^TA$ es invertible.[5] Así pues $\boldsymbol{\lambda} = (A^TA)^{-1}A^T\mathbf{x}$ y dado que $\mathcal{P}_W(\mathbf{x}) = A\boldsymbol{\lambda}$ tenemos una expresión de la proyección que resulta

$$\mathcal{P}_W(\mathbf{x}) = A(A^TA)^{-1}A^T\mathbf{x} \tag{8.4}$$

A la matriz $A(A^TA)^{-1}A^T$ se le conoce como *matriz de proyección*.

Si por el contrario, $A^TA$ no es invertible, entonces hemos de resolver el sistema dado en (8.3) para obtener $\boldsymbol{\lambda}$. ¿Puede el lector averiguar por qué este sistema tiene solución?

**Nota 8.6**

Si el producto escalar no es el habitual, entonces hemos de usar la matriz de Gram, resultando que

$$(\mathbf{x} - A\boldsymbol{\lambda}) \perp W \Rightarrow A^T P_{\mathcal{B}}(\mathbf{x} - A\boldsymbol{\lambda}) = \mathbf{0} \Rightarrow \mathcal{P}_W(\mathbf{x}) = A(A^T P_{\mathcal{B}} A)^{-1} A^T P_{\mathcal{B}}\mathbf{x}$$

**Ejemplo 8.8**

Calculemos la matriz de proyección sobre el subespacio

$$W = L((1, 2, -1), (1, 0, -1))$$

Si consideramos la matriz

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & -1 \end{pmatrix}$$

entonces es evidente que $W = \operatorname{Im}(A)$. Según (8.4), la matrix de la proyección sobre $W$ vendrá dada por

$$A(A^TA)^{-1}A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

[5] De hecho se tiene que $\operatorname{rango}(A) = \operatorname{rango}(A^TA)$. Nótese que $A \in \mathcal{M}_{n\times r}(\mathbb{R})$ con $\operatorname{rango}(A) = r < n$ y $A^TA \in \mathcal{M}_{r\times r}(\mathbb{R})$. Si consideramos $\mathbf{x} \in \ker(A^TA)$ entonces $A^TA\mathbf{x} = \mathbf{0}$, de donde se deduce que $\mathbf{x}^T \cdot (A^TA\mathbf{x}) = 0$, o lo que es igual $(A\mathbf{x})^T \cdot (A\mathbf{x}) = 0$, es decir $A\mathbf{x} = \mathbf{0}$, y por tanto $\mathbf{x} \in \ker(A)$. Con esto se prueba que $\ker(A^TA) \subset \ker(A)$. La inclusión contraria es evidente, y como ambos espacios coinciden, lógicamente también los rangos de ambas matrices.

Es decir, si ahora queremos calcular la proyección del vector $(1, 1, 2)$ sobre $W$, no tenemos más que multiplicar por dicha matriz:

$$\mathcal{P}_W((1,1,2)) = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix}$$

Finalmente damos una propiedad esencial de la proyección que aplicaremos en la siguiente sección.

**Proposición 8.10**

Si $E$ es un e.e., $W$ un subespacio vectorial de $E$ y $\mathbf{x} \in E$ con $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ tal que $\mathbf{x}_1 \in W$, entonces $\|\mathcal{Q}_W(\mathbf{x})\| \leq \|\mathbf{x}_2\|$.

*Demostración:*
De la hipótesis del resultado es evidente que $\mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1$, de modo que

$$\|\mathbf{x}_2\|^2 = \|\mathbf{x} - \mathbf{x}_1\|^2 = \|\mathcal{P}_W(\mathbf{x}) + \mathcal{Q}_W(\mathbf{x}) - \mathbf{x}_1\|^2$$

y usando el Teorema de Pitágoras (pues $\mathcal{P}_W(\mathbf{x}) - \mathbf{x}_1 \in W$ y $\mathcal{Q}_W(\mathbf{x}) \in W^\perp$)

$$= \|\mathcal{P}_W(\mathbf{x}) - \mathbf{x}_1\|^2 + \|\mathcal{Q}_W(\mathbf{x})\|^2 \geq \|\mathcal{Q}_W(\mathbf{x})\|^2$$

■

Es interesante prestar atención a este resultado, pues afirma que dada cualquier descomposición de un vector $\mathbf{x}$ en una suma $\mathbf{x}_1 + \mathbf{x}_2$ donde $\mathbf{x}_1 \in W$, entonces el error entre $\mathbf{x}$ y su proyección ortogonal, esto es el vector $\mathcal{Q}_W(\mathbf{x})$ es siempre más pequeño, en norma, que $\mathbf{x} - \mathbf{x}_1$. La figura 8.3 trata de ilustrar este hecho. Por simplicidad hemos considerado $W$ una recta vectorial (punteada en el dibujo). Las dos primeras figuras muestran descomposiciones del vector $\mathbf{x}$ en suma de un elemento de $W$, $\mathbf{x}_1$, más otro vector $\mathbf{x}_2$. La tercera, muestra la descomposición correspondiente a la proyección ortogonal sobre $W$. Podemos apreciar que las longitudes de los vectores $\mathbf{x}_2$ **siempre** serán mayores que las de $\mathcal{Q}_W(\mathbf{x})$.

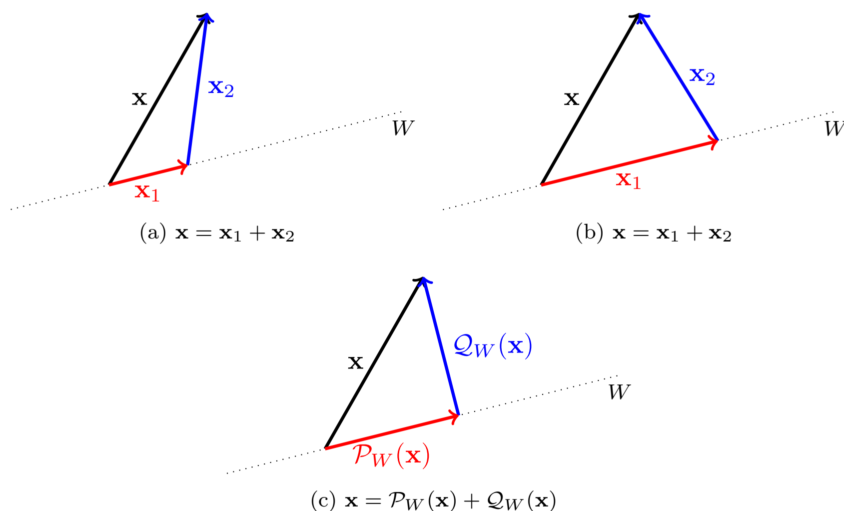



Figura 8.3: Ilustración de la Proposición 8.10

La consecuencia de este resultado es la siguiente: $\mathcal{P}_W(\mathbf{x})$, esto es la proyección ortogonal respecto de un subespacio $W$, es la *mejor aproximación* a $\mathbf{x}$ que podemos encontrar en $W$. Dicho de otra forma, la proyección de un vector $\mathbf{x}$ sobre $W$ es el vector de $W$ que está más cerca a $\mathbf{x}$ de entre todos los que hay en $W$. El concepto de proximidad viene dado en función de la norma que induce el producto escalar en el que estemos trabajando. Por tanto se tiene que

$$\mathcal{P}_W(\mathbf{x}) \text{ es la solución del problema } \min_{\mathbf{y}\in W} \|\mathbf{x}-\mathbf{y}\| \tag{8.5}$$

**Ejemplo 8.9**

Encontrar la mejor aproximación a la función

$$f(x) = \begin{cases} -1 & \text{si } -\pi \le x \le 0, \\ 1 & \text{si } 0 < x \le \pi, \end{cases}$$

en los subespacios $W_k = L(\{\operatorname{sen}(nx), \cos(nx)\})$, $n \le 0, 1, 2, \dots, k$ de $\mathcal{C}([0,\pi])$ con el producto escalar

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

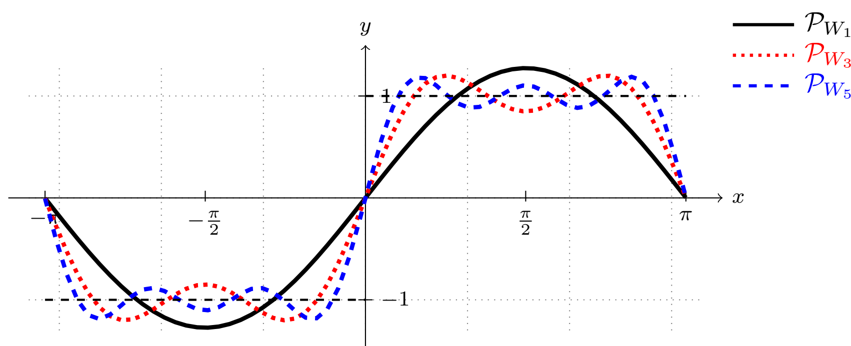



Figura 8.4: Función $f$ y sus proyecciones ortogonales en $W_k$, $k = 1, 3, 5$

Es decir nos están pidiendo la proyección ortogonal de la función $f$ en varios subespacios $W_k$, cada uno de ellos mayor que el anterior.

Puesto que los vectores que generan cada subespacio $W_k$ conforman una base ortogonal (se deja al lector la comprobación), el cálculo de la proyección es inmediato a partir de (II) la nota 8.5. Los coeficientes vendrán dados por

$$\lambda_0 = \frac{\langle f, 1\rangle}{\|1\|^2} = 0, \quad \lambda_n = \frac{\langle f, \operatorname{sen}(nx)\rangle}{\|\operatorname{sen}(nx)\|^2} = \frac{2}{n\pi}(1 - (-1)^n),$$

$$\mu_n = \frac{\langle f, \cos(nx)\rangle}{\|\cos(nx)\|^2} = 0, \quad n = 1, \dots, k$$

Es decir, la proyección en $W_1$, $W_2$ coincide y es $\mathcal{P}_{W_1}(f(x)) = \frac{4}{\pi}\operatorname{sen} x$; para $W_3$ y $W_4$, la proyección es la misma y resulta

$$\mathcal{P}_{W_3}(f(x)) = \frac{4}{\pi}\operatorname{sen} x + \frac{4}{3\pi}\operatorname{sen}(3x)$$

mientras que en $W_5$ y $W_6$ la proyección es

$$\mathcal{P}_{W_3}(f(x)) = \frac{4}{\pi}\operatorname{sen} x + \frac{4}{3\pi}\operatorname{sen}(3x) + \frac{4}{5\pi}\operatorname{sen}(5x)$$

En el espacio generado por la funciones trigonométricas $\{\operatorname{sen}(nx), \cos(mx)\}$, que son ortogonales respecto de este producto escalar, los coeficientes de la proyección son conocidos como *coeficientes de Fourier*.

En la figura 8.4, están representadas la función y sus proyecciones.

8 3

## MÉTODO DE LOS MÍNIMOS CUADRADOS

El método de los mínimos cuadrados es una aplicación de la propiedad descrita en la Proposición 8.10 sobre la proyección ortogonal que permite resolver *de forma aproximada* sistemas de ecuaciones.

Supongamos que tratamos de resolver un sistema de la forma $A\mathbf{x} = \mathbf{b}$, donde $A \in \mathcal{M}_{m\times n}(\mathbb{K})$, $\mathbf{x} \in \mathbb{K}^n$ y $\mathbf{b} \in \mathbb{K}^m$. Identificando la matriz $A$ con la matriz de una determinada aplicación lineal que la define, en una base dada, podemos entender que el sistema tiene solución siempre y cuando $\mathbf{b} \in \mathrm{Im}(A)$. O dicho de otro modo, la columna correspondiente al segundo miembro debe ser linealmente dependiente de las columnas de la matriz $A$, que en esencia es lo que afirma el Teorema de Rouché-Frobenius.

Supongamos que no existe solución para este sistema. En este caso, puede ser interesante estudiar para qué vector o vectores $\tilde{\mathbf{x}}$, $A\tilde{\mathbf{x}} \approx \mathbf{b}$, esto es, tratamos de buscar una *solución aproximada* del sistema.

Una forma de interpretar esta aproximación es la siguiente: si $A\tilde{\mathbf{x}} \approx \mathbf{b}$, entonces podemos intentar encontrar $\tilde{\mathbf{x}}$ tal que $\|A\tilde{\mathbf{x}} - \mathbf{b}\|$ sea lo menor posible, esto es, tratamos de resolver el problema

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|$$

Obsérvese que si el sistema sí tiene solución, entonces el mínimo del problema anterior es cero, y por tanto $A\mathbf{x} = \mathbf{b}$; pero si no hay solución, tiene sentido buscar el vector $\mathbf{x}$ que haga menor esa norma, y por tanto que más se parece a una solución.

Por otra parte, dado cualquier vector $\mathbf{x}$, es evidente que $A\mathbf{x} \in \mathrm{Im}(A)$, luego el problema anterior se puede escribir en los términos

$$\min_{\mathbf{y}\in \mathrm{Im}(A)} \|\mathbf{y} - \mathbf{b}\|$$

Si ahora miramos (8.5), el mínimo anterior se alcanza para $\mathcal{P}_{\mathrm{Im}(A)}(\mathbf{b})$.

Más aun, puesto que la proyección se puede calcular a través de la matriz de proyección obtenida en (8.4), entonces el sistema se transforma en

$$A\tilde{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$$

de aquí se deduce que

$$\tilde{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$$

aunque en realidad no es necesario invertir, sino resolver el sistema

$$A^TA\tilde{\mathbf{x}} = A^T\mathbf{b}$$

**Ejemplo 8.10**

Resolver de forma aproximada el sistema

$$\left.\begin{array}{rcl} x_1 + x_2 + x_3 & = & 1 \\ x_1 + x_2 - x_3 & = & 1 \\ x_3 & = & 1 \end{array}\right\}$$

Está claro que el sistema dado es incompatible. Definiendo la aplicación de matriz

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

el sistema anterior se escribe como $A\mathbf{x} = \mathbf{b}$ con $\mathbf{b} = (1,1,1)$. Para resolverlo no tenemos más que resolver el sistema

$$A^T A\tilde{\mathbf{x}} = A^T \mathbf{b}$$

que en nuestro caso resulta

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

esto es

$$\begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \Leftrightarrow \left\{\begin{array}{rcl} x_1 + x_2 & = & 1 \\ x_3 & = & \frac{1}{3} \end{array}\right.$$

que posee infinitas soluciones, de la forma $(\alpha, 1-\alpha, \frac{1}{3})$. Cualquiera de ellas es una solución aproximada.[6]

## 8 3 1 Ajuste de datos

Un ejemplo típico en el que se emplea el método de los mínimos cuadrados es en el de los problemas de ajuste de datos. Una situación habitual en muchas aplicaciones sucede cuando disponemos de una serie de datos que han sido tomados y queremos encontrar una función que represente de la forma más

[6] En estos casos es frecuente seleccionar una de las soluciones atendiendo a algún criterio como por ejemplo la de menor norma.

precisa posible tal conjunto de datos. La elección del tipo de función a usar depende del fenómeno con el que estemos estudiando, y en la práctica se suelen usar funciones sencillas, como pueden ser rectas o funciones cuadráticas. Veamos un ejemplo.

**Ejemplo 8.11**

Encontrar la recta que mejor aproxima a los puntos $(1, 1)$, $(2, 3)$, $(3, 5)$ y $(4, 6)$.

En este caso buscamos una función lineal, de la forma $y = a+bx$ que se ajuste de forma eficiente a los puntos dados. Si todos los puntos coincidieran sobre la misma recta, está claro que se debería poder resolver el siguiente sistema

$$\left.\begin{array}{rcl} a+b & = & 1 \\ a+2b & = & 3 \\ a+3b & = & 5 \\ a+4b & = & 6 \end{array}\right\} \tag{8.6}$$

Puesto que dicha recta no existe, el sistema no tiene solución y lo que haremos será buscar una solución aproximada usando el método de los mínimos cuadrados. Procediendo como en el ejemplo 8.10, bastará resolver el sistema $A^T A\mathbf{x} = A^T\mathbf{b}$ donde

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 6 \end{pmatrix}$$

En este caso resulta

$$\left.\begin{array}{rcl} 4a+10b & = & 15 \\ 10a+30b & = & 46 \end{array}\right\}$$

cuya solución es $a = -\frac{1}{2}$ y $b = \frac{17}{10}$.

La figura 8.5 muestra los puntos y la recta obtenida, también conocida como *recta de regresión*.

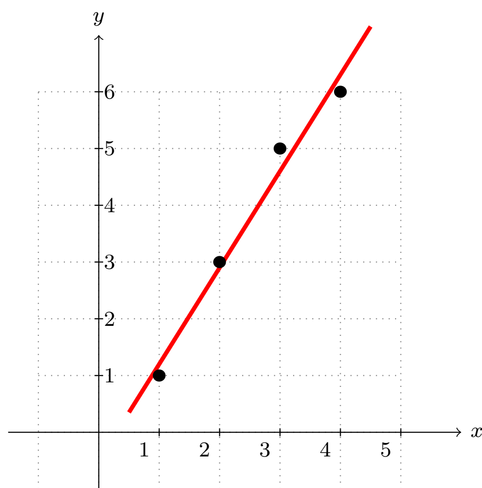



Figura 8.5: Ajuste de datos lineal

## 8 4 CÁLCULO CON PYTHON

La introducción de diversos productos escalares en este tema nos invita a seguir explorando nuevas capacidades de Python, que nos ayudarán a realizar este nuevo tipo de cálculos.

El producto escalar habitual en vectores de $\mathbb{R}^n$ puede ser llevado a cabo desde NumPy con la función `dot`. Por ejemplo

$$(1, 2, 3) \cdot (-1, 0, 2) = 5$$

corresponde a

```
>>> from numpy import array,dot
>>> a=array([1,2,3])
>>> b=array([-1,0,2])
>>> dot(a,b)
5
```

Por el contrario, SymPy no posee ninguna función específica para calcular el producto escalar, por lo que usamos simplemente la expresión $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$, pero sí posee un atributo para calcular la norma inducida por el producto escalar habitual:

```
>>> from sympy import Matrix
>>> a=Matrix([1,-2,-3,0])
>>> b=Matrix([-1,0,1,2])
```

```
>>> a.T*b
[-4]
>>> a.norm()
14**(1/2)
```

es decir, $(1,-2,-3,0)\cdot(-1,0,1,2) = -4$ y $\|(1,-2,-3,0)\| = \sqrt{14}$.

Los productos escalares más laboriosos que se han llevado a cabo son los que involucran el cálculo de integrales. Tanto NumPy como SymPy permiten realizar cálculo integral, aunque el primero lo hace mediante técnicas numéricas, mientras que el segundo permite hacer cálculo simbólico, adaptándose mejor a nuestras necesidades.

```
>>> from sympy import symbols,integrate
>>> t=symbols('t')
>>> p=1+t+t**2
>>> q=2-3*t+t**2
>>> integrate(p*q,(t,-1,1))
22/5
```

esto es

$$\int_{-1}^{1}(1+t+t^2)(2-3t+t^2)\,dt = \frac{22}{5}$$

De este modo es posible calcular las normas y la ortogonalidad de las funciones del ejemplo 8.9:

```
>>> from sympy import symbols,integrate,cos,sin,pi
>>> x,n,m=symbols('x,n,m')
>>> integrate(sin(n*x)*cos(m*x),(x,-pi,pi))
0
>>> integrate(sin(n*x)*sin(m*x),(x,-pi,pi))
2*m**3*cos(pi*m)*sin(pi*n)/(2*m**2*n**2 - m**4 - n**4) + 2*n
    **3*cos(pi*n)*sin(pi*m)/(2*m**2*n**2 - m**4 - n**4) - 2*m
    *n**2*cos(pi*m)*sin(pi*n)/(2*m**2*n**2 - m**4 - n**4) -
    2*n*m**2*cos(pi*n)*sin(pi*m)/(2*m**2*n**2 - m**4 - n**4)
>>> integrate(sin(n*x)**2,(x,-pi,pi))
(pi*n/2 - cos(pi*n)*sin(pi*n)/2)/n - (cos(pi*n)*sin(pi*n)/2
    - pi*n/2)/n
>>> simplify(_)
(pi*n - cos(pi*n)*sin(pi*n))/n
```

Si leemos con calma el resultado se observa que:

$$\int_{-\pi}^{\pi}\operatorname{sen}(nx)\cos(mx)\,dx = 0$$

$$\int_{-\pi}^{\pi}\operatorname{sen}(nx)\operatorname{sen}(mx)\,dx = \frac{2}{2m^2n^2-m^4-n^4}\Big(m^3\cos(\pi m)\operatorname{sen}(\pi n)+$$
$$n^3\cos(\pi n)\operatorname{sen}(\pi m) - mn^2\cos(\pi m)\operatorname{sen}(\pi n) - nm^2\cos(\pi n)\operatorname{sen}(\pi m)\Big) = 0$$

$$\int_{-\pi}^{\pi} \operatorname{sen}(nx)^2\,dx = \frac{1}{n}\Big(\pi n - \cos(\pi n)\sin(\pi n)\Big) = \pi$$

donde hemos de tener en cuenta que si $n$ y $m$ son números naturales, entonces $\operatorname{sen}(n\pi) = 0$, $\forall n \in \mathbb{N}$.

La solución aproximada de sistemas mediante el método de mínimos cuadrados está implementada en NumPy con una función que empleamos en el tema 4 para obtener el rango de una matriz. Dicha función, `linalg.lstsq`, proporciona la solución mediante mínimos cuadrados de un sistema de la forma $A\mathbf{x} = \mathbf{b}$, así como información adicional sobre el sistema, como el rango de la matriz. Esta función tiene dos parámetros de entrada, la matriz $A$ y el segundo miembro $\mathbf{b}$. Por ejemplo, si queremos resolver el sistema (8.6) del ejemplo 8.11, haremos lo siguiente

```
>>> from numpy import array,linalg
>>> a=array([[1,1],[1,2],[1,3],[1,4]])
>>> b=array([1,3,5,6])
>>> linalg.lstsq(a,b)
(array([-0.5,  1.7]), array([ 0.3]), 2, array([ 5.77937881,
     0.77380911]))
```

Observamos que la respuesta de la función `lstsq`$(A, \mathbf{b})$ es amplia: el primer argumento es un arreglo con la solución aproximada del sistema; el segundo argumento es el *residuo*, que es igual a $\|A\mathbf{x} - \mathbf{b}\|$, con $\mathbf{x}$ la solución aproximada, salvo que el rango de la matriz no sea máximo, en cuyo caso es un arreglo vacío; el tercer argumento es, como ya vimos, el rango de la matriz $A$, mientras que el último proporciona los *valores singulares* de la matriz $A$.[7]

Si resolvemos un sistema cuya matriz no tiene rango máximo, y que por tanto conduce a infinitas soluciones, como en el ejemplo 8.10,

```
>>> from numpy import array, linalg
>>> a=array([[1,1,1],[1,1,-1],[0,0,1]])
>>> b=array([1,1,1])
>>> linalg.lstsq(a,b)
(array([ 0.5       ,  0.5       ,  0.33333333]), array([],
    dtype=float64), 2, array([  2.00000000e+00,   1.73205081e
    +00,   7.23911034e-17]))
```

vemos que la solución obtenida es la de menor norma.

## 8.5 APLICACIÓN: LIGHTS OUT!, SEGUNDA PARTE

En la sección 4.7 presentamos el juego *Lights Out!* y lo planteamos como un problema algebraico equivalente a resolver un sistema de ecuaciones. En concreto, vimos que una determinada configuración de luces encendidas y apagadas

[7]Los valores singulares de una matriz son las raíces de los autovalores de la matriz $A^T A$.

representada por un vector $\mathbf{b} \in \mathbb{Z}_2^{25}$ es alcanzable si resulta ser una combinación lineal de las columnas de la matriz $A$ definida en (4.13).

Ahora vamos a usar lo aprendido sobre aplicaciones lineales y ortogonalidad para identificar de forma sencilla cuándo un vector $\mathbf{b}$ corresponde a una configuración alcanzable. Para ello hemos de prestar atención a un resultado, cuya demostración está propuesta en el ejercicio 31, y que afirma que si $A$ es la matriz de un endomorfismo, entonces

$$\ker(A) = \operatorname{Im}(A^T)^\perp \quad \text{y} \quad \operatorname{Im}(A) \perp \ker(A^T)^\perp$$

En nuestro caso particular, la matriz $A$ de (4.13) es simétrica, por lo que el resultado anterior se traduce en

$$\ker(A)^\perp = \operatorname{Im}(A)$$

Dicho de otro modo, un vector que pertenezca a $\operatorname{Im}(A)$ debe ser, necesariamente, ortogonal a $\ker(A)$.

En resumen, si un vector $\mathbf{b} \in \mathbb{Z}_2^{25}$ es alcanzable, entonces debe pertenecer al subespacio generado por las columnas de $A$, es decir, debe pertenecer al subespacio $\operatorname{Im}(A)$, y según acabamos de ver, debe ser ortogonal a cualquier vector de $\ker(A)$. Tenemos por tanto un sencillo test para comprobar si $\mathbf{b}$ es alcanzable.

Vamos pues a determinar el espacio $\ker(A)$: para ello hemos de resolver el sistema

$$\begin{aligned} x_1 + x_2 + x_6 &= 0 \\ x_1 + x_2 + x_3 + x_7 &= 0 \\ x_2 + x_3 + x_4 + x_8 &= 0 \\ &\vdots \\ x_{19} + x_{23} + x_{24} + x_{25} &= 0 \\ x_{20} + x_{24} + x_{25} &= 0 \end{aligned}$$

obteniéndose que una base de soluciones del mismo viene dada por los vectores

$$\begin{aligned} \mathbf{n}_1 &= (0,1,1,1,0,1,0,1,0,1,1,1,0,1,1,1,0,1,0,1,0,1,1,1,0) \\ \mathbf{n}_2 &= (1,0,1,0,1,1,0,1,0,1,0,0,0,0,0,1,0,1,0,1,1,0,1,0,1) \end{aligned}$$

Por ejemplo, supongamos que la configuración inicial es la que aparece en la figura 8.6. Es decir,

$$\mathbf{b} = (1,1,1,0,1,0,1,0,1,1,0,1,0,0,1,1,0,1,0,1,1,0,1,1,1)$$

Ver si es posible apagar todas las luces de la figura 8.6 equivale a comprobar si

$$\mathbf{b} \cdot \mathbf{n}_1 = \mathbf{b} \cdot \mathbf{n}_2 = 0$$

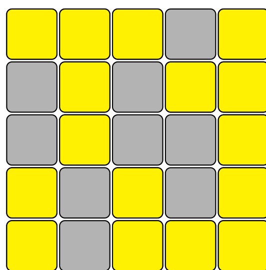


Figura 8.6: ¿Se pueden apagar todas las luces?

Haciendo el producto escalar habitual, módulo 2, se verifica que, en efecto, $\mathbf{b} \in \ker(A)^{\perp}$.

Con un poco de paciencia se puede usar el método de Gauss para el sistema $A\mathbf{x} = \mathbf{b}$, realizando todas las operaciones módulo 2, para obtener una solución:

$$\mathbf{x} = (0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,0,0,1)$$

Pero además, puesto que $A\mathbf{n}_1 = \mathbf{0}$ y $A\mathbf{n}_2 = \mathbf{0}$, se tendrá que

$$\mathbf{x},\ \mathbf{x} + \mathbf{n}_1,\ \mathbf{x} + \mathbf{n}_2,\ \mathbf{x} + \mathbf{n}_1 + \mathbf{n}_2$$

también son soluciones.

## 8 6 EJERCICIOS

### Ejercicios de repaso

**E.1** Sea $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + (x_1 + x_2)(y_1 + y_2) + (x_1 + x_2 + x_3)(y_1 + y_2 + y_3)$ un producto escalar en $\mathbb{R}^3$. Encontrar la matriz del producto escalar respecto de la base canónica de $\mathbb{R}^3$.

**E.2** Ortogonalizar la base de $\mathbb{R}^3$ dada por los vectores $\mathbf{x}_1 = (1,0,0)$, $\mathbf{x}_2 = (4,-2,0)$ y $\mathbf{x}_3 = (1,1,5)$.

**E.3** Encontrar una base ortogonal del subespacio vectorial

$$M = \{\mathbf{x} \in \mathbb{R}^3 : x_1 + 2x_2 + 3x_3 = 0\}$$

**E.4** Decidir si los siguientes pares de subespacios son ortogonales:

(a) $L_1 = L((2,1,-1),(0,1,1))$, $L_2 = L((-1,2,0))$

(b) $L_1 = L((-3,0,1)$, $L_2 = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : 3x_1 - x_3 = 0\}$

(c) $L_1 = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1 + x_2 - x_3 = 0,\ x_1 - x_2 = 0\}$,
$L_2 = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1 + x_2 = 0\}$

**E.5** Se considera el espacio de polinomios $\mathbb{P}^3_{\mathbb{R}}[-1,1]$ con el producto escalar

$$\langle p(x), q(x)\rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

y sea $W$ el subespacio generado por los polinomios $x-1$ y $x^2$. Determinar una base ortogonal de $W^\perp$.

**E.6** Sean los vectores de $\mathbb{R}^5$, $\mathbf{a} = (1,2,0,-1,0)$, $\mathbf{b} = (0,1,-1,1,0)$ y $\mathbf{c} = (0,0,1,2,1)$. Descomponer $\mathbf{c}$ en dos vectores, uno de ellos combinación lineal de $\mathbf{a}$ y $\mathbf{b}$ y el otro ortogonal a éste.

**E.7** ¿Qué combinación lineal de los vectores $(1,2,-1)$ y $(1,0,1)$ es la más cercana a $(2,1,1)$?

**E.8** En el espacio vectorial

$$V = \{f : [0,1] \to \mathbb{R} : f \text{ es derivable, } f(0) = 0\},$$

se considera el producto escalar

$$\langle f, g\rangle = \int_0^1 f'(x)g'(x)\,dx$$

(a) Ortogonaliza las funciones $f_1(x) = x$, $f_2(x) = x^2$.

(b) Halla la proyección ortogonal de $f(x) = e^{-x} - 1$ sobre el subespacio generado por $f_1$ y $f_2$.

**E.9** Para los siguientes apartados, hallar la solución del sistema $A\mathbf{x} = \mathbf{b}$ por mínimos cuadrados:

$$\text{(a) } A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}$$

$$\text{(b) } A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 4 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

### Problemas

**E.10** Sean $\mathbf{u}_1 = (-2,-1,1)$, $\mathbf{u}_2 = (0,-1,0)$ y $\mathbf{u}_3 = (1,-1,0)$ vectores l.i. de $\mathbb{R}^3$. Definimos un producto escalar en $\mathbb{R}^3$ afirmando que $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ es una base ortonormal. Encontrar la expresión analítica de este producto escalar en la base canónica de $\mathbb{R}^3$.

**E.11** Ortogonalizar los vectores de $\mathbb{R}^4$: $\mathbf{x}_1 = (1,1,0,0)$, $\mathbf{x}_2 = (1,2,0,0)$ y $\mathbf{x}_3 = (0,0,-1,1)$, y hallar una base ortogonal de $\mathbb{R}^4$ que los contenga.

**E.12** Encontrar una base ortonormal para los siguientes subespacios vectoriales:

(a) El subespacio de $\mathbb{P}^3_{\mathbb{R}}[-1,1]$

$$M = \{p(x) \in \mathbb{P}^3_{\mathbb{R}}[-1,1] : xp'(x) = p(x)\}$$

con el producto escalar del ejercicio 5.

(b) El subespacio de $\mathcal{M}_2(\mathbb{R})$

$$N = \{A \in \mathcal{M}_2(\mathbb{R}) : \operatorname{tr}(A) = 0\}$$

con el producto escalar definido en el ejercicio 23.

**E.13** Encontrar el complemento ortogonal de los subespacios vectoriales $M$ y $N$ del ejercicio 12.

**E.14** En $\mathcal{M}_2(\mathbb{R})$ con el producto escalar del ejercicio 23, hallar el complemento ortogonal del subespacio de las matrices diagonales de $\mathcal{M}_2(\mathbb{R})$ y el complemento ortogonal de las matrices simétricas de $\mathcal{M}_2(\mathbb{R})$.

**E.15** Determinar la proyección ortogonal y la mínima distancia de una matriz $X \in \mathcal{M}_2(\mathbb{R})$ al subespacio de las matrices diagonales, esto es, se trata de calcular $\min\limits_{D\in\mathcal{D}} \|D - X\|$, donde $\mathcal{D}$ es el conjunto de las matrices diagonales de $\mathcal{M}_2(\mathbb{R})$.

**E.16** Considera $P_1$ la matriz de la aplicación que a cada vector le asocia su proyección sobre el espacio generado por el vector $\mathbf{a}_1$, donde $\mathbf{a}_1 = (-1,2,2)$. Haz lo mismo con $P_2$, donde $\mathbf{a}_2 = (2,2,-1)$. Multiplica las matrices resultantes $P_1P_2$ y explica por qué sale ese resultado.

**E.17** Sea $C([-\pi,\pi])$ el espacio euclídeo de las funciones continuas en el intervalo $[-\pi,\pi]$, con el producto escalar

$$\langle f(t), g(t)\rangle = \int_{-\pi}^{\pi} f(t)g(t)\,dt$$

Sea $L$ el subespacio vectorial de $V$ generado por los vectores 1, $\operatorname{sen} t$ y $\cos t$. Se pide

(a) Encontrar la proyección ortogonal del vector $h(t) = (t-\pi)$ sobre $L$.

(b) Hallar el vector $p(t) \in L$ que haga mínimo el valor de la integral

$$\int_{-\pi}^{\pi} [(t-\pi) - p(t)]^2\,dt$$

(Ayuda: $\displaystyle\int_{-\pi}^{\pi} \sin^2 t\,dt = \int_{-\pi}^{\pi} \cos^2 t\,dt = \pi$).

*E.18 Hallar el polinomio $Q(t) = t^3 + at^2 + bt + c$ para el cual la integral

$$\int_{-1}^{1} (Q(t))^2 \, dt$$

sea lo más pequeña posible.

E.19 En un concesionario de automóviles se han realizado una serie de observaciones tratando de relacionar el precio $y_i$ de los vehículos con su peso $x_i$. Dados los datos de la tabla adjunta hallar por el método de los mínimos cuadrados el mejor ajuste lineal $y = ax + b$.

| peso (en Tm) | 0.8 | 1 | 1.2 | 1.3 |
|---|---|---|---|---|
| precio (en millones) | 1 | 2 | 3 | 5 |

E.20 Una editorial analiza las ventas realizadas de los libros de uno de sus mejores autores durante los últimos cuatro años, reflejadas en la tabla

| año | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| venta (en millones) | 1 | 1.5 | 3 | 5 |

(a) Hallar el ajuste lineal por el método de los mínimos cuadrados.

(b) ¿Cuántos libros de dicho autor espera vender la editorial en el quinto año?

E.21 Se consideran los puntos $(1, 1)$, $(2, 4)$, $(3, 7)$, $(4, 9)$. Ajustar dichos puntos

(a) Por una función polinómica de segundo grado $y = a + bx + cx^2$ utilizando el método de los mínimos cuadrados.

(b) Por una función $y = a + bx + cx^2 + dx^3$.

### 8.6.1 Ejercicios teóricos

E.22 Decir cuáles de las siguientes funciones $\langle \cdot\,, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ definen un producto escalar en $\mathbb{R}^2$:

(a) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 2x_2y_2 + 2x_2y_1 - 3x_1y_2$

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 - x_2y_2$

(c) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_2 + 2x_1y_2 + 3y_1x_2 + 7x_2y_1$

(d) $\langle \mathbf{x}, \mathbf{y} \rangle = 2x_1y_2 + 3x_2y_2 + 2x_1y_2 + 2x_2y_1$

E.23 En el espacio vectorial de las matrices $\mathcal{M}_2(\mathbb{R})$, demostrar que el producto

$$\langle A, B \rangle = \operatorname{tr}(A\,B^T)$$

es un producto escalar.

E.24 Sea $V$ un espacio euclídeo. Demostrar:

(a) $2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right) = \|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2$

(b) $4\langle \mathbf{u}, \mathbf{v}\rangle = \|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2$

(c) $2\langle \mathbf{u}, \mathbf{v}\rangle = \|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2$

E.25 Demostrar que la desigualdad de Schwarz es una igualdad si y sólo si los vectores son proporcionales.

E.26 Si $\mathbf{x}$ e $\mathbf{y}$ son dos vectores cualesquiera de un e.e. $E$, demostrar que

$$\|\mathbf{x}-\mathbf{y}\| \geq \|\mathbf{x}\| - \|\mathbf{y}\|$$

E.27 Sea E un e.e., demostrar que $\mathbf{u}+\mathbf{v}$ y $\mathbf{u}-\mathbf{v}$ son ortogonales si y sólo si $\|\mathbf{u}\| = \|\mathbf{v}\|$.

E.28 Demostrar la siguiente generalización del teorema de Pitágoras:

$$\|\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_n\|^2 = \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \cdots + \|\mathbf{x}_n\|^2$$

si los vectores $\mathbf{x}_1, \ldots, \mathbf{x}_n$ son ortogonales entre sí.

*E.29 Sea $V$ un e.e. y $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ un conjunto de vectores ortogonales. Demostrar que:

(a) $\displaystyle\sum_{i=1}^{n} \frac{|\langle \mathbf{v}, \mathbf{u}_i\rangle|^2}{\|\mathbf{u}_i\|^2} \leq \|\mathbf{v}\|^2, \ \forall \mathbf{v} \in V$ (desigualdad de Bessel).

(b) Si los $\mathbf{u}_i$ forman una base ortogonal de $V$ entonces

$$\langle \mathbf{v}, \mathbf{w}\rangle = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{u}_i\rangle\langle \mathbf{u}_i, \mathbf{w}\rangle}{\|\mathbf{u}_i\|^2}, \quad \forall \mathbf{v}, \mathbf{w} \in V \qquad \text{(identidad de Parserval)}$$

E.30 Demostrar que si $\langle \mathbf{x}, \mathbf{x}\rangle = \displaystyle\sum_{i=1}^{n} x_i^2$, con $(x_1, \ldots, x_n)$ las coordenadas de $\mathbf{x}$ en una base $\mathcal{B}$, entonces esta base es ortonormal.

*E.31 Sea $A$ la matriz de un endomorfismo de $\mathbb{R}^n$. Demostrar que

$$\ker(A) = \operatorname{Im}(A^T)^\perp \ \text{y} \ \operatorname{Im}(A) = \ker(A^T)^\perp$$

## Ejercicios adicionales

E.32 Usar Python para comprobar que el conjunto de funciones de $\mathcal{C}([0, 2\pi])$:

$$\left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\operatorname{sen} t, \frac{1}{\sqrt{\pi}}\cos t, \frac{1}{\sqrt{\pi}}\operatorname{sen} 2t, \frac{1}{\sqrt{\pi}}\cos 2t, \ldots, \frac{1}{\sqrt{\pi}}\operatorname{sen} nt, \frac{1}{\sqrt{\pi}}\cos nt,\right\}$$

es un conjunto ortonormal con el producto escalar

$$\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\,dt$$

E.33 Considérese el juego descrito en el ejercicio 31 del tema 4. Encontrar el espacio $\ker(A)$, donde $A$ es la matriz del sistema que define el juego y razonar si toda configuración puede ser alcanzable para el mismo.

# 9 Espacio afín

En los temas anteriores hemos interpretado en más de una ocasión diversos elementos que aparecen en los espacios vectoriales como conjuntos geométricos tales como rectas o planos. Sin embargo, desde un punto de riguroso, esta interpretación no es exacta pues los objetos geométricos están formados por puntos y éstos son objetos fijos, mientras que el concepto de vector es una magnitud que no está sujeta a una posición espacial determinada. En definitiva, los espacios vectoriales no son los conjuntos apropiados para describir los objetos geométricos que el lector habrá estudiado en cursos anteriores.

En este tema vamos a introducir el espacio afín como el marco geométrico adecuado en el que poder trabajar con los objetos típicamente estudiados en geometría plana o espacial, y en el que poder plantear problemas métricos y transformaciones lineales.

9 1

## ESPACIO AFÍN Y ESPACIO MÉTRICO

La definición de espacio afín[1] recupera el concepto intuitivo de vector como el objeto que marca la dirección entre dos puntos del espacio, asociándolo a un elemento de un espacio vectorial.

**Definición 9.1**

Sea $V$ un espacio vectorial y $\mathcal{A}$ un conjunto no vacío cuyos elementos denominaremos *puntos*. Se dice que $\mathcal{A}$ es un *espacio afín* asociado a $V$ si existe una aplicación

$$\begin{array}{rccc} \varphi: & \mathcal{A}\times\mathcal{A} & \longrightarrow & V \\ & (P,Q) & \longrightarrow & \varphi(P,Q)=\overrightarrow{PQ} \end{array}$$

que posee las siguientes propiedades:

(I) Para todo punto $M\in\mathcal{A}$ y todo vector $\mathbf{x}\in V$ existe un único punto $N\in\mathcal{A}$ tal que $\overrightarrow{MN}=\mathbf{x}$.

[1] El término fue introducido por Euler en 1748 su libro *Introductio in analysin infinitorum*.

(II) Para cualesquiera tres puntos $M$, $N$ y $P$ de $\mathcal{A}$ se tiene que $\overrightarrow{MN} + \overrightarrow{NP} = \overrightarrow{MP}$.

Un vector $\mathbf{x} = \overrightarrow{PQ}$ diremos que tiene *origen* en $P$ y *extremo* en $Q$. También se define la *dimensión* de un espacio afín como la dimensión del espacio vectorial asociado. En todo lo que sigue trataremos solo con espacios afines de dimensión finita.

En definitiva, un espacio afín es una colección de puntos de manera que para cada par ordenado de ellos, $P$ y $Q$, podemos construir el vector con origen $P$ y extremo $Q$, esto es $\overrightarrow{PQ}$. Se verifican las siguientes propiedades:

**Proposición 9.1**

Dados los puntos $M$, $N$, $P$ y $Q$, se tiene:

(I) Si $\overrightarrow{MN} = \overrightarrow{PQ}$ entonces $\overrightarrow{MP} = \overrightarrow{NQ}$.

(II) $\overrightarrow{MN} = \mathbf{0}$ si y solo si $M = N$.

(III) $\overrightarrow{MN} = -\overrightarrow{NM}$.

*Demostración:*

(I) Sumando a ambos miembros de la igualdad $\overrightarrow{NP}$,

$$\overrightarrow{MN} + \overrightarrow{NP} = \overrightarrow{NP} + \overrightarrow{PQ}$$

y usando la propiedad (II) de la definición, $\overrightarrow{MP} = \overrightarrow{NQ}$.

(II) Nuevamente por la propiedad (II) de la definición,

$$\overrightarrow{MN} + \overrightarrow{NN} = \overrightarrow{MN} \Rightarrow \overrightarrow{NN} = \mathbf{0}$$

Recíprocamente, si $\overrightarrow{MN} = \overrightarrow{NN} = \mathbf{0}$, por la propiedad (I) de la definición $M = N$.

(III) Es evidente que $\overrightarrow{MN} + \overrightarrow{NM} = \overrightarrow{MM} = \mathbf{0}$, luego $\overrightarrow{MN} = -\overrightarrow{NM}$.

■

### 9.1.1 Sistema de referencia y coordenadas

Si en el espacio afín $n$-dimensional $\mathcal{A}$ se considera un punto fijo $O \in \mathcal{A}$, que se denomina *origen de coordenadas*, y una base $\mathcal{B}$ del espacio vectorial asociado, el conjunto $\mathcal{R} = \{O; \mathcal{B}\}$ es un *sistema de referencia* que nos permite obtener las *coordenadas* de un punto cualquiera $P \in \mathcal{A}$, a través de las coordenadas del vector $\overrightarrow{OP}$ respecto de la base $\mathcal{B}$. Se notará $P(p_1, \ldots, p_n)_{\mathcal{R}}$, o simplemente $P(p_1, \ldots, p_n)$ cuando no haya posibilidad de confusión.[2]

Si tenemos dos puntos $P(p_1, \ldots, p_n)_{\mathcal{R}}$ y $Q(q_1, \ldots, q_n)_{\mathcal{R}}$, entonces el vector $\overrightarrow{PQ}$ satisface

$$\overrightarrow{PQ} = \overrightarrow{PO} + \overrightarrow{OQ} = \overrightarrow{OQ} - \overrightarrow{OP}$$

y dado que las coordenadas de $\overrightarrow{OP}$ y $\overrightarrow{OQ}$ coinciden con las de $P$ y $Q$ respectivamente, entonces

$$\overrightarrow{PQ} = (q_1 - p_1, \ldots, q_n - p_n)_{\mathcal{B}}$$

que coincide con la expresión que el lector habrá visto en cursos previos de geometría.

Cuando el punto $O(0, \ldots, 0)$ y $\mathcal{B}$ es la base canónica, lo denominaremos *sistema de referencia cartesiano.*

### 9.1.2 Cambio de sistema de referencia

Al igual que ocurre con las bases, dados dos sistemas de referencia $\mathcal{R} = \{O; \mathcal{B}\}$ y $\mathcal{R}' = \{O'; \mathcal{B}'\}$ en un espacio afín $\mathcal{A}$, buscamos la relación existente entre las coordenadas en cada uno de los sistemas de referencia de un mismo punto $P \in \mathcal{A}$.

Supongamos que $P$ tiene coordenadas $P(x_1, \ldots, x_n)_{\mathcal{R}} = (x'_1, \ldots x'_n)_{\mathcal{R}'}$ y pongamos que $O'(c_1, \ldots, c_n)_{\mathcal{R}}$. Dado que para cada punto $P$ se verifica que

$$\overrightarrow{OP} = \overrightarrow{OO'} + \overrightarrow{O'P}$$

entonces, si $M_{\mathcal{B}'}^{\mathcal{B}} = (a_{ij})$ es la matriz de cambio de base de $\mathcal{B}'$ a $\mathcal{B}$, se tiene que

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} + M_{\mathcal{B}'}^{\mathcal{B}} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

que son las **ecuaciones del cambio de sistema de referencia**. Estas ecua-

[2]Nótese la omisión del signo de igualdad en la notación de los puntos, no así en la de los vectores.

ciones se pueden abreviar del siguiente modo:

$$\begin{pmatrix} 1 \\ \hline x_1 \\ \vdots \\ x_n \end{pmatrix} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_n & a_{n1} & \cdots & a_{nn} \end{array}\right) \begin{pmatrix} 1 \\ \hline x'_1 \\ \vdots \\ x'_n \end{pmatrix} \tag{9.1}$$

Estas ecuaciones permiten calcular el cambio de sistema de referencia inverso usando para ello la inversa de esta matriz, tal y como sucede con los cambios de base en espacios vectoriales.

### 9 1 3 Espacio afín euclídeo

Si el espacio vectorial asociado a un espacio afín $\mathcal{A}$ es además un espacio euclídeo, es decir, dotado de un producto escalar, se dirá que $\mathcal{A}$ es un *espacio afín euclídeo*. Se dirá que un sistema de referencia en un espacio afín $\mathcal{R} = \{O; \mathcal{B}\}$ es *ortonormal* si la base $\mathcal{B}$ es ortonormal.

**Definición 9.2**

Dado un conjunto cualquiera $A$, una aplicación:

$$\begin{array}{rccc} d: & A \times A & \longrightarrow & \mathbb{R}^+ \cup \{0\} \\ & (x,y) & \longrightarrow & d(x,y) \end{array}$$

que verifique:

(I) $d(x,y) = 0 \Leftrightarrow x = y$, $\forall x, y \in A$

(II) $d(x,y) = d(y,x)$, $\forall x, y \in A$.

(III) $d(x,y) \leq d(x,z) + d(z,y)$, $\forall x, y, z \in A$.

se dice que es una *distancia*. En tal caso decimos que $A$ es un *espacio métrico*.

En un espacio afín euclídeo $\mathcal{A}$ podemos definir una distancia $d(P,Q) = \|\overrightarrow{PQ}\|$, de modo que $\mathcal{A}$ es un espacio métrico.

**Ejemplo 9.1**

El plano $\mathbb{R}^2$, entendido como un conjunto de puntos, tiene estructura de espacio afín, si lo consideramos conjuntamente con el espacio vectorial $\mathbb{R}^2$. Si

tomamos el sistema de referencia formado por el punto $O(0,0)$ y la base canónica de $\mathbb{R}^2$, $\{(1,0),(0,1)\}$ tenemos el sistema de referencia canónico en el que las coordenadas de cada punto $P(x,y)$ coinciden con las coordenadas del vector que une el origen $O$ con el propio punto $P$.

Si en este espacio consideramos el producto escalar habitual, la métrica a la que hace referencia la definición anterior corresponde a la distancia *euclídea* que el lector ya conoce. Esto es, la distancia entre dos puntos $P(x,y)$ y $Q(x',y')$ viene dada por

$$d(P,Q) = \|\overrightarrow{PQ}\| = \sqrt{(x'-x)^2 + (y'-y)^2}$$

9 2

## VARIEDADES AFINES

**Definición 9.3**

Dado un espacio afín $\mathcal{A}$ asociado a un espacio vectorial $V$, se dice que un subconjunto $L \subset \mathcal{A}$ es una *variedad afín* de $\mathcal{A}$ si existe un punto $P \in L$ tal que el conjunto

$$W = \{\overrightarrow{PQ} : Q \in L\}$$

es un subespacio vectorial de $V$.

En definitiva, una variedad afín es a un espacio afín, lo que un subespacio vectorial es a un espacio vectorial. Las variedades afines también se puede definir del siguiente modo: dado $P \in \mathcal{A}$ y $W$ un subespacio vectorial, una variedad afín $L$ es un subconjunto de $\mathcal{A}$ de la forma

$$L = \{Q \in \mathcal{A} : \overrightarrow{PQ} \in W\}$$

De este modo, podemos escribir una variedad afín como $L = P + W = \{P + \mathbf{w} : \mathbf{w} \in W\}$. Al subespacio $W$ se le denomina *dirección* de la variedad.

**Proposición 9.2**

La variedad afín es independiente del punto usado en su definición. Es decir, si $L = P + W$ y $P' \in L$, entonces $L = P' + W$.

*Demostración:*
Observemos en primer lugar que si $P' \in L$ entonces $\overrightarrow{PP'} = \mathbf{w} \in W$. Sea ahora $Q \in P + W$, entonces $Q = P + \mathbf{x}$ para cierto $\mathbf{x} \in W$. Así,

$$\overrightarrow{P'Q} = \overrightarrow{P'P} + \overrightarrow{PQ} = -\mathbf{w} + \mathbf{x} \in W$$

luego $Q \in P' + W$; es decir, $P + W \subset P' + W$. La inclusión contraria se prueba de idéntica forma.

■

**Proposición 9.3**

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{K})$ y $\mathbf{b} \in \mathbb{K}^m$ y consideremos el sistema de ecuaciones lineal $A\mathbf{x} = \mathbf{b}$. Si el conjunto de soluciones de este sistema es no vacío y $P$ es una solución del mismo, entonces dicho conjunto define una variedad lineal en el espacio afín que pasa por $P$ y tiene como dirección el subespacio vectorial formado por las soluciones del sistema homogéneo asociado, $A\mathbf{x} = \mathbf{0}$.

Recíprocamente, toda variedad afín puede definirse como el conjunto de soluciones de un sistema lineal de ecuaciones $A\mathbf{x} = \mathbf{b}$.

La demostración es consecuencia inmediata de la Proposición 3.2 que trata de la estructura de las soluciones de un sistema no homogéneo, y de los Teoremas 4.7 y 4.8.

**Ejemplo 9.2**

Veamos algunas construcciones con las que el lector estará familiarizado:

(I) En el espacio afín $\mathbb{R}^2$, consideramos el punto $P(1, 2)$ y el subespacio $W = L((1, 1))$. La variedad afín $L = P + W$ corresponde a la recta que pasa por $P$ y tiene la dirección $W$. Los puntos de esta recta tendrán por expresión

$$(x, y) = (1, 2) + t(1, 1),\ t \in \mathbb{R} \Rightarrow \begin{cases} x = 1 + t \\ y = 2 + t \end{cases}$$

que corresponden a la *ecuación vectorial* y *ecuación paramétrica* de la recta, respectivamente. Si despejamos el parámetro $t$ de las ecuaciones paramétricas se obtiene

$$x - 1 = y - 2 \Rightarrow x - y = -1$$

que es la *ecuación general* o *cartesiana* de la recta.

Es posible obtener directamente la ecuación general de la recta si obtenemos primero un sistema de ecuaciones homogéneo para la dirección de ésta, en este caso, la ecuación $x - y = 0$. Por tanto, según la Proposición 9.3, la ecuación será un sistema no homogéneo con la misma matriz, es decir de la forma $x - y = \alpha$. Para determinar el parámetro $\alpha$ basta imponer que el punto $P(1, 2)$ verifique dicha ecuación, luego $\alpha = -1$.

(II) En $\mathbb{R}^3$ consideramos un punto $P(3, -1, 1)$ y el subespacio $W$ generado por el vector $\mathbf{v} = (2, 0, 1)$. Para obtener las ecuaciones de la recta correspondiente a la variedad afín $L = P + W$ procedemos como en el ejemplo anterior, obteniendo un sistema de ecuaciones homogéneo del espacio $W$:

$$\text{rango}\begin{pmatrix} 2 & 0 & 1 \\ x_1 & x_2 & x_3 \end{pmatrix} = 1 \Rightarrow \begin{cases} x_2 = 0 \\ 2x_3 - x_1 = 0 \end{cases}$$

de manera que las ecuaciones de la recta serán

$$\begin{cases} x_2 = \alpha \\ 2x_3 - x_1 = \beta \end{cases} \quad \Rightarrow \alpha = -1,\ \beta = -1$$

donde $\alpha$ y $\beta$ se obtienen al imponer que el punto $P$ satisfaga ambas ecuaciones.

(III) La variedad afín correspondiente al plano que pasa por el punto $P(-1, 2, 0)$ y tiene como vector normal el vector $(1, 1, -1)$ se obtiene considerando el subespacio vectorial ortogonal a dicho vector, que tiene como ecuación $x_1 + x_2 - x_3 = 0$; de modo que la ecuación del plano será

$$x_1 + x_2 - x_3 = \alpha \Rightarrow \alpha = 1$$

donde $\alpha$ es obtenida como en los ejemplos anteriores.

(IV) Para calcular la ecuación del plano que pasa por tres puntos dados, $M(1, 1, 0)$, $N(-1, 2, 1)$ y $P(0, -1, 2)$, nos basta considerar cualquiera de los tres puntos, junto con el subespacio generado por dos vectores independientes del plano, que se obtiene mediante pares de puntos, es decir,

$$M(1, 1, 0); \quad \overrightarrow{MN} = (-2, 1, 1); \quad \overrightarrow{MP} = (-1, -2, 2)$$

En lugar de proceder como en los ejemplos anteriores calculando la ecuación homogénea correspondiente al subespacio vectorial y luego obteniendo el término independiente, podemos también razonar del siguiente modo: cualquier punto del plano buscado $Q(x, y, z)$, satisfará que el vector $\overrightarrow{MQ}$

debe ser linealmente dependiente de los otros vectores $\overrightarrow{MN}$ y $\overrightarrow{MQ}$, de modo que

$$\text{rango}\begin{pmatrix} x-1 & y-1 & z \\ -2 & 1 & 1 \\ -1 & -2 & 2 \end{pmatrix} = 2 \Rightarrow \begin{vmatrix} x-1 & y-1 & z \\ -2 & 1 & 1 \\ -1 & -2 & 2 \end{vmatrix} = 0$$

Resolviendo el determinante se obtiene la ecuación $4x + 3y + 5z = 7$.

### 9.2.1 Posición relativa de variedades afines

**Definición 9.4**

Dadas dos variedades afines $L_1$ y $L_2$, diremos que éstas se *cortan* si $L_1 \cap L_2 \neq \emptyset$. Si las variedades no se cortan diremos que son *paralelas* si $W_1 \subset W_2$ (ó $W_2 \subset W_1$), donde $W_i$ es la dirección de la variedad $L_i$. En caso contrario diremos que las variedades se *cruzan*. Del mismo modo, dos variedades son *perpendiculares* si $W_1 \subset W_2^\perp$ (ó $W_2 \subset W_1^\perp$).

Para estudiar la posición relativa de variedades lineales hemos de estudiar su intersección, resolviendo los sistemas lineales de ecuaciones que las definen. Veamos algunos ejemplos.

**Ejemplo 9.3**

(I) Dada la recta $r \equiv P + \langle \mathbf{v} \rangle$, con $P(-2, 0, -1)$ y $\mathbf{v} = (3, 1, 1)$ y el plano $\pi$ de ecuación $x_1 - x_3 = 2$, para averiguar su posición relativa consideraremos el sistema formado por las ecuaciones de ambas variedades.

Puesto que las ecuaciones de $r$ son $x_1 - 3x_2 = 1$, $x_2 - x_3 = 1$, debemos resolver el sistema

$$\begin{cases} x_1 - 3x_2 = 1 \\ x_2 - 3x_3 = 1 \\ x_1 - x_3 = 2 \end{cases} \Rightarrow x_1 = \tfrac{7}{4},\ x_2 = \tfrac{1}{4},\ x_3 = -\tfrac{1}{4}$$

es decir, la recta y el plano se cortan en el punto $\left(\frac{7}{4}, \frac{1}{4}, -\frac{1}{4}\right)$.

(II) Estudiamos ahora la posición relativa de los planos

$$\pi_1 \equiv x_1 - x_2 + 3x_3 = -1, \quad \pi_2 \equiv x_2 - x_3 = 5$$

Estudiando el sistema formado por ambas ecuaciones resulta un sistema compatible indeterminado, de modo que existen infinitas soluciones, lo cual expresa que ambos planos se cortan en una recta. Para averiguar un punto y un vector que generan dicha recta nos basta resolver el sistema en forma paramétrica, esto es

$$\begin{cases} x_1 - x_2 + 3x_3 = -1 \\ x_2 - x_3 = 5 \end{cases} \Rightarrow (4-2t, 5+t, t) = (4,5,0) + t(-2,1,1)$$

es decir, es la recta generada por el punto $(4,5,0)$ y el vector $(-2,1,1)$.

También es posible obtener la ecuación de la recta resolviendo el sistema homogéneo asociado, con lo que obtendríamos un vector de la misma, obteniéndose luego una solución particular para el sistema.

(III) Posición relativa de las rectas

$$r_1 \equiv \begin{cases} x - y = 1 \\ y + z = 0 \end{cases} \qquad r_2 \equiv \begin{cases} x = 3 \\ y - z = -1 \end{cases}$$

Si tratamos de resolver el sistema formado por las cuatro ecuaciones que definen ambas rectas observamos que éste es incompatible, de manera que las rectas son paralelas o se cruzan. Veamos entonces cómo son sus direcciones: para ello resolvemos los sistemas homogéneos asociados obteniendo:

$$W_{r_1} = L(1,1,-1), \qquad W_{r_2} = L((0,1,1))$$

Puesto que $W_{r_1} \neq W_{r_2}$ las rectas no son paralelas y por tanto se cruzan. De hecho podemos observar que las rectas son perpendiculares pues $W_{r_1} \perp W_{r_2}$.

9 3

## PROBLEMAS MÉTRICOS

El concepto de proyección ortogonal en un espacio euclídeo puede ser extendido a los espacios afines a través del siguiente resultado.

**Proposición 9.4**

Sea $L = A+W$ una variedad afín de un espacio afín euclídeo $\mathcal{A}$ tal que $L \neq \mathcal{A}$ y sea $B \in \mathcal{A}$ tal que $B \notin L$. Entonces existe una variedad unidimensional (una recta) $r \equiv B + V$ tal que $r$ es perpendicular a $L$ y $r \cap L \neq \emptyset$.

*Demostración:*
Consideremos el vector $\overrightarrow{AB}$ y su proyección ortogonal sobre el espacio $W$, esto es $\mathcal{P}_W(\overrightarrow{AB})$. Sea $B' = A + \mathcal{P}_W(\overrightarrow{AB})$, es decir $\overrightarrow{AB'} = \mathcal{P}_W(\overrightarrow{AB})$. Está claro que

$$\overrightarrow{AB} = \overrightarrow{AB'} + \overrightarrow{B'B}$$

y de la definición de proyección ortogonal (Definición 8.8) se tiene que

$$\overrightarrow{AB} = \mathcal{P}_W(\overrightarrow{AB}) + \mathcal{Q}_W(\overrightarrow{AB})$$

Dado que esta descomposición es única se tiene que $Q_W(\overrightarrow{AB}) = \overrightarrow{B'B}$.

Consideramos ahora el subespacio $V$ generado por el vector $\overrightarrow{BB'}$ y la variedad (la recta) $r \equiv B + V$. Está claro que $V \subset W^\perp$, es decir $r$ es perpendicular a $L$ y $B' \in r \cap L$.

■

El punto $B'$ obtenido en la construcción anterior se denomina *proyección ortogonal* del punto $B$ sobre $L$ y se notará por $\mathcal{P}_L(B)$. Es importante resaltar que $B'$ también se obtiene como la intersección entre $L$ y $L_B^\perp = B + W^\perp$.

**Ejemplo 9.4**

Proyectar el punto $A(1, 0, 2)$ sobre la recta $L = P + \langle \mathbf{w} \rangle$, con $P(1, 1, 1)$ y $\mathbf{w} = (-1, 2, 1)$.

Tal y como hemos comentado, para construir $\mathcal{P}_L(A)$ construimos en primer lugar la variedad $L_A^\perp = A + W^\perp$ (donde $W = \langle \mathbf{w} \rangle$), es decir la variedad afín que pasa por $A$ y tiene como dirección el espacio ortogonal al generado por el vector $\mathbf{w}$. Como hemos visto en el tema anterior, un sistema de ecuaciones implícitas para $W^\perp$ viene dado por $-x_1 + 2x_2 + x_3 = 0$, de modo que la variedad $L_A^\perp$ tiene por ecuaciones

$$-x_1 + 2x_2 + x_3 = 1$$

Intersecamos ahora con la recta $L$ de ecuaciones

$$x_1 + x_3 = 2, \;\; x_1 + 2x_2 = 3$$

que da lugar a un sistema cuya solución es la proyección buscada $(\frac{4}{3}, \frac{5}{6}, \frac{2}{3})$.

Como vimos en la sección 9.1, la distancia entre dos puntos en el espacio afín se obtiene a través de la norma del vector que une los puntos. Gracias a la proyección ortogonal podemos obtener la distancia entre un punto y una variedad afín; dado un punto $P$ y una variedad $L$, la distancia de $P$ a $L$, definida por

$$d(P, L) = \min_{Q \in L} \|\overrightarrow{PQ}\|$$

la obtenemos a través del siguiente resultado.

**Proposición 9.5**

Dado $P$ un punto y $L$ una variedad afín de un espacio $\mathcal{A}$, se tiene que $d(P, L) = d(P, \mathcal{P}_L(P))$

*Demostración:*
Sea $L = A + W$, con $A \in \mathcal{A}$ y $W$ la dirección de $L$. Sabemos que $P' = \mathcal{P}_L(P)$ corresponde al punto tal que $\overrightarrow{AP'} = \mathcal{P}_W(\overrightarrow{AP})$. Sea ahora $Q \in L$ cualquiera, entonces

$$\overrightarrow{AP} = \overrightarrow{AQ} + \overrightarrow{QP}$$

y como $\overrightarrow{AQ} \in W$, de la Proposición 8.10 sabemos que

$$\|\overrightarrow{PP'}\| \leq \|\overrightarrow{QP}\|, \quad \forall Q \in L$$

lo que prueba que la menor distancia entre $P$ y cualquier punto de $L$ se obtiene a través de la proyección ortogonal.

■

**Ejemplo 9.5**

En $\mathbb{R}^3$ consideremos un punto $P(x_0, y_0, z_0)$ y un plano de ecuación $ax + by + cz + d = 0$. Para calcular la distancia del punto al plano construimos la variedad ortogonal al plano que pasa por $P$, esto es, $P + \mathbf{w}$ con $\mathbf{w} = (a, b, c)$. Usando la ecuación vectorial de esta recta, para intersecarla con el plano se ha de encontrar $t \in \mathbb{R}$ tal que

$$a(x_0 + at) + b(y_0 + tb) + c(z_0 + tc) + d = 0 \Rightarrow t = -\frac{ax_0 + by_0 + cz_0 + d}{a^2 + b^2 + c^2}$$

Para este valor de $t$ se obtiene el punto $P'$ correspondiente a la proyección del $P$ sobre el plano. Por tanto, la distancia será la norma de este vector:

$$\|\overrightarrow{PP'}\| = \|(ta, tb, tc)\| = \sqrt{t^2(a^2+b^2+c^2)}$$
$$= \left|\frac{ax_0+by_0+cz_0+d}{a^2+b^2+c^2}\right| \sqrt{a^2+b^2+c^2} = \frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}$$

con lo que se obtiene la conocida fórmula de la distancia de un punto a un plano.

También podemos definir la distancia entre dos variedades $L_1$ y $L_2$ como la menor distancia existente entre cualesquiera puntos de las variedades, es decir,

$$d(L_1, L_2) = \min_{\substack{P\in L_1\\ Q\in L_2}} d(P, Q)$$

**Proposición 9.6**

Dadas dos variedades $L_1 = A_1 + W_1$ y $L_2 = A_2 + W_2$ se tiene que

$$d(L_1, L_2) = d(A_1, L)$$

donde $L = A_2 + (W_1 + W_2)$, es decir, la variedad que contiene a $L_2$ y es paralela a $L_1$.

*Demostración:*
Consideremos dos puntos $P_1$ y $P_2$ de $L_1$ y $L_2$, respectivamente. Sea ahora $\mathbf{w}_1 = \overrightarrow{P_1A_1} \in W_1$ y sea $P = P_2 + \mathbf{w}_1$. Nótese que $P \in L$.

Se tiene que

$$\overrightarrow{P_1P_2} = \overrightarrow{P_1A_1} + \overrightarrow{A_1P} + \overrightarrow{PP_2} = \mathbf{w}_1 + \overrightarrow{A_1P} - \mathbf{w}_1 = \overrightarrow{A_1P}$$

de modo que $d(P_1, P_2) = d(A_1, P)$. Puesto que $P \in L$, deducimos que $d(P_1, P_2) \geq d(A_1, L)$. Como esto es cierto cualesquiera que sean los puntos $P_1$ y $P_2$ en $L_1$ y $L_2$, respectivamente, en particular es cierto para los puntos en los que se alcanza la distancia mínima, es decir,

$$d(L_1, L_2) \geq d(A_1, L)$$

Por otro lado, consideremos ahora $Q = \mathcal{P}_L(A_1)$. Puesto que $Q \in L$, podemos escribir $Q = A_2 + \mathbf{v}_1 + \mathbf{v}_2$, para ciertos vectores $\mathbf{v}_1 \in W_1$ y $\mathbf{v}_2 \in W_2$. Si ahora llamamos $P_1 = A_1 - \mathbf{v}_1$ y $P_2 = A_2 + \mathbf{v}_2$, que obviamente verifican que $P_1 \in L_1$ y $P_2 \in L_2$, también se tiene que

$$\overrightarrow{A_1Q} = \overrightarrow{A_1P_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2Q} = -\mathbf{v}_1 + \overrightarrow{P_1P_2} + \mathbf{v}_1 = \overrightarrow{P_1P_2}$$

Como $Q$ es el punto que da la distancia mínima de $A_1$ a $L$, tenemos que $d(A_1, L) = \|\overrightarrow{P_1P_2}\|$. Dado que $P_1$ y $P_2$ están en $L_1$ y $L_2$, respectivamente está claro que

$$d(A_1, L) \geq d(L_1, L_2)$$

De la doble desigualdad se obtiene el resultado.

■

**Ejemplo 9.6**

Calcular la distancia entre las rectas $r$ y $s$ dadas por

$$r \equiv (1, 0, -1) + \langle(0, 1, 1)\rangle, \qquad s \equiv \begin{cases} x - y + z = 1 \\ x + y = 2 \end{cases}$$

Según la proposición anterior, la distancia $d(r, s)$ viene dada por la distancia de un punto de $r$, a la variedad $L$ que contiene a $s$ y es paralela a $r$. Por tanto, construyamos en primer lugar dicha variedad $L$.

Calculamos la dirección de $s$ encontrando una base del subespacio vectorial asociado, esto es

$$\begin{cases} x - y + z = 0 \\ x + y = 0 \end{cases} \quad \Rightarrow \text{Dirección de } s = (-1, 1, 2)$$

Consideremos un punto cualquiera de $s$, por ejemplo $(1, 1, 1)$, y construyamos la variedad $L$, en este caso el plano, que pasa por $(1, 1, 1)$ y tiene dirección $\langle(0, 1, 1), (-1, 1, 2)\rangle$. Como vimos en (IV) del ejemplo 9.2, la ecuación del plano vendrá dada resolviendo

$$\begin{vmatrix} x - 1 & y - 1 & z - 1 \\ 0 & 1 & 1 \\ -1 & 1 & 2 \end{vmatrix} = 0 \Rightarrow x - y + z = 1$$

Luego $d(r, s) = d((1, 0, -1), L)$, que según hemos visto en el ejemplo 9.5 viene dada por

$$d\left((1, 0, -1), L\right) = \frac{|-1|}{\sqrt{3}}$$

**Ejemplo 9.7**

Obtener la perpendicular común a las rectas

$$r \equiv \begin{cases} 2x + y = 1 \\ x + z = -1 \end{cases} \qquad s \equiv \begin{cases} x - 2y - z = 1 \\ x - y = 0 \end{cases}$$

En primer lugar consideremos $W$, la dirección perpendicular a $r$ y $s$ simultáneamente: para ello, obtenemos primero las direcciones de $r$ y $s$:

$$\begin{cases} 2x + y = 0 \\ x + z = 0 \end{cases} \Rightarrow (1, -2, -1) \qquad \begin{cases} x - 2y - z = 0 \\ x - y = 0 \end{cases} \Rightarrow (1, 1, -1)$$

y la dirección perpendicular a $r$ y $s$ vendrá dada por las soluciones del subespacio

$$\begin{cases} x - 2y - z = 0 \\ x + y - z = 0 \end{cases} \Rightarrow W = \langle (1, 0, 1) \rangle$$

Ahora, la perpendicular común se obtiene mediante la intersección de la variedad $\pi_1$ que contiene a $r$ y es paralela a $W$ con la variedad $\pi_2$ que contiene a $s$ y es paralela a $W$:

$$\pi_1 \equiv \begin{vmatrix} x & y-1 & z+1 \\ 1 & -2 & -1 \\ 1 & 0 & 1 \end{vmatrix} = 0 \Rightarrow x + y - z = 2$$

$$\pi_2 \equiv \begin{vmatrix} x & y & z+1 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{vmatrix} = 0 \Rightarrow x - 2y - z = 1$$

(nótese que el punto $(0, 1, -1) \in r$ y $(0, 0, -1) \in s$). Entonces, la perpendicular común es la recta

$$\begin{cases} x + y - z = 2 \\ x - 2y - z = 1 \end{cases}$$

## 9.4 APLICACIONES AFINES

Las aplicaciones afines son similares a las aplicaciones lineales pero definidas en espacios afines. Dado que una aplicación afín transforma unos puntos en otros, podemos considerar una aplicación asociada que transforma los vectores

que conectan esos puntos en otros vectores. En esencia, una aplicación afín será aquella que hace que la aplicación asociada sea lineal.

**Definición 9.5**

Sean $\mathcal{A}$ y $\mathcal{A}'$ dos espacios afines. Una aplicación $f : \mathcal{A} \to \mathcal{A}'$ se dice que es una *aplicación afín* si existe un punto $O \in \mathcal{A}$ tal que la aplicación lineal $\overrightarrow{f} : V \to V'$, donde $V$ y $V'$ son los espacios vectoriales asociados a los espacios afines $\mathcal{A}$ y $\mathcal{A}'$, respectivamente, verifica $\overrightarrow{f}(\overrightarrow{OX}) = \overrightarrow{f(O)f(X)}$, $\forall X \in \mathcal{A}$.

La aplicación $\overrightarrow{f}$ se dirá la aplicación lineal asociada a $f$.

El siguiente resultado muestra que la aplicación lineal asociada no depende de la elección del punto $O$.

**Proposición 9.7**

Si $f$ es una aplicación afín, entonces la aplicación lineal asociada verifica $\overrightarrow{f}(\overrightarrow{AB}) = \overrightarrow{f(A)f(B)}$, $\forall A, B \in \mathcal{A}$.

*Demostración:*
Dado que $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$ y que $\overrightarrow{f}$ es lineal,

$$\overrightarrow{f}(\overrightarrow{AB}) = \overrightarrow{f}(\overrightarrow{OB}) - \overrightarrow{f}(\overrightarrow{OA}) = \overrightarrow{f(O)f(B)} - \overrightarrow{f(O)f(A)} = \overrightarrow{f(A)f(B)}$$

■

De la definición de aplicación afín se sigue que ésta sólo depende de la imagen de un punto $O$ dado y de la aplicación lineal asociada, pues de la expresión $\overrightarrow{f}(\overrightarrow{OX}) = \overrightarrow{f(O)f(X)}$, se sigue que

$$f(X) = f(O) + \overrightarrow{f}(\overrightarrow{OX}) \tag{9.2}$$

## 9.4.1 Matriz de una aplicación afín

Consideremos dos sistemas de referencia $\mathcal{R} = \{O; \mathcal{B}\}$ y $\mathcal{R}' = \{O'; \mathcal{B}'\}$ de los espacios $\mathcal{A}$ y $\mathcal{A}'$, respectivamente. De la expresión (9.2), conocidas las coordenadas del punto $f(O)(c_1, \dots, c_m)_{\mathcal{R}'}$ y la matriz de la aplicación lineal $\overrightarrow{f}$ en las bases $\mathcal{B}$ y $\mathcal{B}'$, $M = (a_{ij})$, se tiene que si $X(x_1, \dots, x_n)_{\mathcal{R}}$ y

$f(X)(y_1, \ldots, y_m)_{\mathcal{R}'}$, entonces

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} + \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

que en formato abreviado escribiremos como $f(X) = f(O) + MX$. La expresión matricial anterior puede ser escrita del siguiente modo

$$\left(\begin{array}{c} 1 \\ \hline y_1 \\ \vdots \\ y_m \end{array}\right) = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline c_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_m & a_{m1} & \cdots & a_{mn} \end{array}\right) \left(\begin{array}{c} 1 \\ \hline x_1 \\ \vdots \\ x_n \end{array}\right)$$

donde la matriz

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline C & M \end{array}\right)$$

se denomina *matriz asociada a la aplicación* $f$ respecto de los sistemas de referencia $\mathcal{R}$ y $\mathcal{R}'$.

**Ejemplo 9.8**

Dado un vector fijo $\mathbf{v}$, consideremos la aplicación afín $f : \mathbb{R}^n \to \mathbb{R}^n$ definida por

$$f(P) = P + \mathbf{v}$$

es decir, la aplicación afín cuya aplicación lineal asociada es la identidad. Esta aplicación afín es una *traslación* de vector $\mathbf{v}$, es decir, a cada punto le asocia el trasladado por el vector $\mathbf{v}$.

Veamos cuál es la aplicación lineal asociada $\overrightarrow{f}$. Dado dos puntos $P$ y $Q$, al trasladarlos mediante el vector $\mathbf{v}$ se tiene que $f(P) = P + \mathbf{v}$ y $f(Q) = Q + \mathbf{v}$. Por tanto

$$\overrightarrow{f}(\overrightarrow{PQ}) = \overrightarrow{f(P)f(Q)} = \overrightarrow{PQ}$$

de manera que $\overrightarrow{f}$ es la aplicación identidad. Además, la imagen del origen $O$ es

$$f(O) = O + \mathbf{v}$$

así pues, la matriz de esta aplicación es

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{v} & I \end{array}\right)$$

**Ejemplo 9.9**

Encontrar la ecuación de la simetría respecto del plano de ecuación

$$x + y - z = 1$$

En primer lugar calcularemos la matriz de la aplicación lineal asociada. Para ello, consideramos la simetría correspondiente a la dirección de la variedad afín dada, esto es, el subespacio $x_1 + x_2 - x_3 = 0$. Para calcular la matriz de esta simetría procedemos como en el ejemplo 5.5, obteniendo en primer lugar una matriz respecto de una base adecuada y posteriormente realizando un cambio de base. Así, la matriz de la aplicación lineal en la base canónica es

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

lo que significa que la matriz de la aplicación afín que buscamos será de la forma:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ c_1 & \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ c_2 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ c_3 & \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ z \end{pmatrix}$$

y para calcular los coeficientes $c_1$, $c_2$ y $c_3$ bastará conocer la imagen de cualquier punto de la aplicación; por ejemplo, como cualquier punto del plano $x+y-z=1$ es fijo, se tendrá que $f(1,0,0) = (1,0,0)$. Imponiendo esta condición resulta que

$$c_1 + \tfrac{1}{3} = 1, \quad c_2 - \tfrac{2}{3} = 0, \quad c_3 + \tfrac{2}{3} = 0$$

es decir, la simetría buscada tiene por ecuaciones

$$f(x,y,z) = \tfrac{1}{3}\left(2 + x - 2y + 2z, 2 - 2x + y + 2z, -2 + 2x + 2y + z\right)$$

## 9.5 APLICACIÓN: MOVIMIENTOS RÍGIDOS

Intuitivamente, el concepto de movimiento es el de una aplicación que "mueve" los objetos en el plano o el espacio sin modificar su forma. En el contexto de las aplicaciones afines, los movimientos se definen de forma clara y precisa a través de sus aplicaciones lineales asociadas. Un *movimiento rígido* es

una aplicación afín cuya aplicación lineal asociada es lo que se denomina una *isometría* o *aplicación ortogonal*. Las aplicaciones ortogonales son aquellas que conservan el producto escalar, es decir

$$\langle \mathbf{u}, \mathbf{v}\rangle = \langle \overrightarrow{f}(\mathbf{u}, \overrightarrow{f}(\mathbf{v})\rangle$$

lo cual se traduce fácilmente (véase ejercicio 19) en conservación de distancias, esto es:

$$\|\overrightarrow{PQ}\| = \|\overrightarrow{f}(\overrightarrow{PQ})\|$$

Desde un punto de vista intuitivo, está claro que una aplicación afín asociada a una isometría hará que los objetos se transformen sin variar las distancias entre sus puntos, lo cual se traduce lógicamente en que la aplicación lo único que hace es modificar la situación espacial del objeto.

El nombre de aplicación ortogonal para denominar a este tipo de aplicaciones lineales no es casual. Resulta que si consideramos una base ortonormal, la matriz de una aplicación ortogonal respecto de esa base resulta ser una matriz ortogonal, esto es, $A^{-1} = A^T$ (cf. Definición 8.5). Tales matrices tienen propiedades muy interesantes: por ejemplo, es muy fácil comprobar que si $A$ es la matriz de una isometría, entonces $|A| = \pm 1$; y aun más, los autovalores de una isometría tienen siempre módulo (o valor absoluto, en caso de que sean reales) igual a 1.

De este modo, es sencillo identificar todas las posibles isometrías tanto en el plano como en el espacio. Por ejemplo, en $\mathbb{R}^2$, como los autovalores tiene que ser 1, $-1$ o un complejo de módulo 1, el cual puede escribirse en la forma $\cos\alpha + i \operatorname{sen}\alpha$, entonces la forma de Jordan (real) de una isometría tiene que corresponder a alguna de las siguientes:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{pmatrix}$$

Por tanto, las aplicaciones afines que son movimientos en el plano tendrán que tener como aplicación lineal asociada alguna de los cuatro tipos anteriores.

Para caracterizar las aplicaciones afines que corresponden a movimientos hemos de prestar atención a sus *puntos fijos*. Un aplicación afín $f : \mathcal{A} \to \mathcal{A}$ tiene un punto fijo $P$ si $f(P) = P$. Por (9.2), $f(P) = f(O) + MP$, donde $M$ es la matriz de la aplicación lineal asociada. Está claro entonces que en un punto fijo se debe cumplir $(M - I_2)P = -f(O)$, donde $I_2$ es la matriz identidad. Este sistema tendrá solución, según el Teorema de Rouché-Frobenius, si

$$\operatorname{rango}(M - I_2) = \operatorname{rango}(M - I_2 | - f(O))$$

Nótese entonces que si $\operatorname{rango}(M - I_2) = 2$, el sistema tendrá siempre un punto fijo, lo que se puede dar en los casos

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{pmatrix}$$

El único moviento plano con un punto fijo es un *giro* de centro el punto fijo y ángulo $\alpha$ (figura 9.1a). Nótese que si $\alpha = \pi$, ambas matrices coinciden.

Si rango$(M - I_2) = 1$, lo cual corresponde necesariamente a que la isometría tenga como forma de Jordan

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

entonces, si hay puntos fijos, entonces debe haber infinitos, generando una variedad lineal de dimensión uno, esto es, una recta. El único movimiento del plano con una recta de puntos fijos es la *simetría* de eje dicha recta (figura 9.1b). También podría ocurrir que para rango$(M - I_2) = 1$ no existieran puntos fijos (aunque la isometría siga siendo la misma); en este caso, lo que tenemos es una *simetría con deslizamiento* (figura 9.1c).

Por último, si la forma de Jordan de la isometría corresponde a la matriz identidad, entonces está claro que rango$(M - I_2) = 0$, y aquí caben dos posibilidades: o bien todos los puntos son fijos, es decir, el movimiento es la identidad, o bien no existe ningún punto fijo, lo que significa que el movimiento es una *traslación* (figura 9.1d).

## 9 6 CÁLCULO CON PYTHON

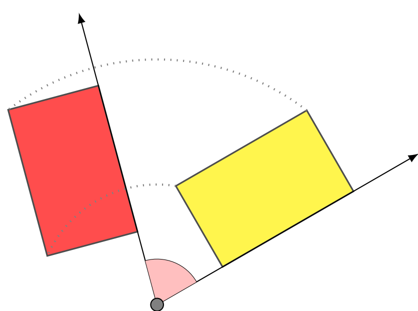
Aunque los cálculos involucrados en este tema no requieren de la introducción de nuevas herramientas de Python, mostraremos en esta sección nuevas instrucciones para manejar matrices que nos ayudarán a resolver algunos ejemplos sencillos.

Comencemos calculando las coordenadas de un mismo punto en dos sistemas de referencia distintos: dados $\mathcal{R} = \{O; \mathcal{B}\}$ el sistema de referencia cartesiano y $\mathcal{R}' = \{O'; \mathcal{B}'\}$, donde

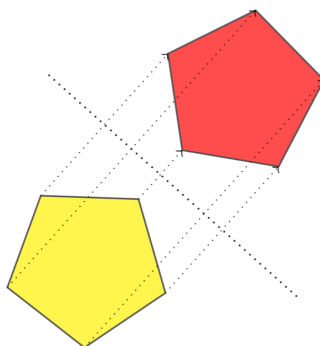
$$O'(1, 1, 0), \; \mathcal{B}' = \{(1, 0, -1), (1, 1, 0), (0, 0, 1)\}$$

para calcular las coordenadas del punto $(1, 2, 3)_{\mathcal{R}}$ respecto de $\mathcal{R}'$ hemos de construir la matriz del cambio de base de $\mathcal{R}'$ a $\mathcal{R}$ de (9.1):

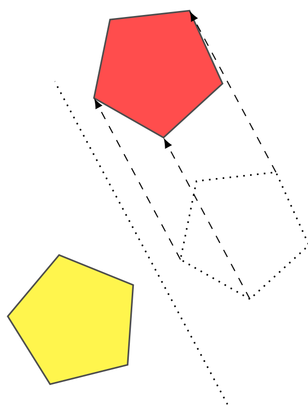
```
>>> from numpy import matrix,concatenate,zeros,bmat,r_,dot
>>> A=matrix('1 0 -1; 1 1 0; 0 0 1').T
>>> M=concatenate((zeros((1,3)),A))
>>> M
matrix([[ 0.,  0.,  0.],
        [ 1.,  1.,  0.],
        [ 0.,  1.,  0.],
        [-1.,  0.,  1.]])
>>> a=concatenate((bmat(r_[1,1,1,0]).T,M),1)
>>> a
```



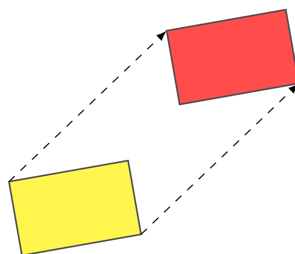
(a) Rotación



(b) Simetría



(c) Simetría con deslizamiento



(d) Traslación

Figura 9.1: Movimientos en el plano

```
11 matrix([[ 1.,  0.,  0.,  0.],
12         [ 1.,  1.,  1.,  0.],
13         [ 1.,  0.,  1.,  0.],
14         [ 0., -1.,  0.,  1.]])
```

En este caso hemos construido la matriz de cambio de sistema de referencia usando nuevas instrucciones de NumPy, como `concatenate`, que se usa para unir dos matrices. En la línea 3 adjuntamos una fila de ceros a la matriz $M$ del cambio de base (formada por los vectores de $\mathcal{B}'$ por columnas), mientras que en la línea 9, hemos adjuntado a la matriz `M` una columna formada por un 1 y las coordenadas de $O'$. Nótese que hemos usado la función `r_` seguida de una lista para crear un arreglo; es decir, la orden `r_[1,1,1,0]` es equivalente a escribir `array([1,1,1,0])`. Luego convertimos este arreglo en matriz con la función `bmat`. Obsérvese también que para adjuntar una columna, en lugar de una fila, usamos un segundo parámetro opcional en la instrucción `concatenate`.

Para transformar coordenadas nos bastará multiplicar la matriz de cambio por el vector $(1, 1, 2, 3)$

```
15 >>> dot(a,r_[1,1,2,3])
16 matrix([[ 1.,  4.,  3.,  2.]])
```

es decir, $(1, 2, 3)_{\mathcal{R}} = (4, 3, 2)_{\mathcal{R}'}$.

Estas instrucciones para manejar matrices están disponibles solo en NumPy. Para realizar algo similar con SymPy hemos de usar las funciones `row_join` y `col_join`. Por ejemplo, podemos construir la matriz del ejemplo 9.8 del siguiente modo:

```
>>> from sympy import Matrix,eye,zeros
>>> from numpy import r_
>>> f=Matrix(r_[1]).row_join(zeros((1,3)))
>>> v=Matrix(r_[1,3,2]).T
>>> A=f.col_join(v.row_join(eye(3)))
>>> A
[1, 0, 0, 0]
[1, 1, 0, 0]
[3, 0, 1, 0]
[2, 0, 0, 1]
```

Tambien podemos aprovechar el manejo de expresiones simbólicas con el módulo SymPy para obtener la ecuación de un plano que pasa por los puntos $(1, 0, 0)$, $(2, 0, -1)$ y $(1, 1, 2)$, según hemos visto en (iv) del Ejemplo 9.2,

```
>>> from sympy import symbols,Matrix
>>> x,y,z=symbols('xyz')
>>> A=Matrix([[x-1,y,z],[1,0,-1],[0,1,2]])
>>> A
[-1 + x, y,  z]
```

```
[      1,  0,  -1]
[      0,  1,   2]
>>> A.det()
-1 + x + z - 2*y
```

9 7

## EJERCICIOS

### Ejercicios de repaso

**E.1** Se considera en un espacio afín $\mathcal{A}$, el sistema de referencia cartesiano $\mathcal{R} = \{O; \{\mathbf{u}_1, \mathbf{u}_2\}\}$ y otro sistema de referencia $\mathcal{R}' = \{O'; \{\mathbf{v}_1, \mathbf{v}_2\}\}$, con $0'(1,2)$, $\mathbf{v}_1 = (2,2)$ y $\mathbf{v}_2 = (-1,2)$. Encontrar las ecuaciones del cambio de sistema de referencia y la ecuación de la recta $r \equiv x + y - 5 = 0$ respecto de $\mathcal{R}'$.

**E.2** En los siguientes apartados encontrar las ecuaciones de la recta:

(a) Que pasa por los puntos $P(1,2,-1)$ y $Q(2,1,3)$.

(b) Pasa por el punto $Q(0,1,-1,0)$ y tiene como dirección el subespacio generado por el vector $(0,1,-1,1)$.

(c) Paralela a la recta de ecuaciones $x - y + z = 1$, $2x + y = 0$ y que pasa por el punto $P(0,0,0)$.

**E.3** Estudiar la posición relativa de las rectas de ecuaciones

$$r \equiv (2,1,4) + \langle(3,2,-1)\rangle \qquad s \equiv \begin{cases} x - 3y + z = -2 \\ 4x + 2y - 3z = -1 \end{cases}$$

**E.4** En los siguientes apartados hallar la ecuación del plano que:

(a) Pasa por los puntos $A(1,0,0)$, $B(2,0,-1)$ y $C(1,1,2)$.

(b) Pasa por el punto $P(-1,0,2)$ y es paralelo a las rectas del ejercicio 3.

(c) Pasa por el origen y es perpendicular a la recta

$$r \equiv \begin{cases} x + y - 2z = 1 \\ x - y + 3z = 0 \end{cases}$$

**E.5** Halla la distancia del punto $P(1,3,-1)$ a la recta de ecuaciones paramétricas

$$r \equiv \begin{cases} x = 2t \\ y = 2 - t \\ z = 1 + 2t \end{cases}$$

y al plano $\pi \equiv x + y + 2z = 0$.

**E.6** Halla la distancia entre las variedades afines

$$r_1 \equiv \begin{cases} x + 2z + t = 0 \\ x - y - 2t = 1 \\ x - t = 2 \end{cases} \qquad r_2 \equiv (0,3,2,1) + \langle(0,1,0,1)\rangle$$

**E.7** Encuentra las ecuaciones de las siguientes transformaciones afines de $\mathbb{R}^3$:

(a) Simetría respecto del plano $x + y + z = 1$.

(b) Simetría respecto de la recta de ecuaciones $x = y = z$.

(c) Simetría respecto del punto $(1,0,1)$.

**E.8** En los siguientes apartados calcular la proyección ortogonal del punto $P$ sobre la variedad $L$, donde

(a) $P(0,1,-2)$, $L \equiv x + y + z = 1$

(b) $P(2,0,-1)$, $L \equiv \begin{cases} 2x + y - z = 0 \\ x - z = 2 \end{cases}$

(c) $P(0,1,-1,2)$, $L \equiv 2x + y + z - t = 2$

## Problemas

**E.9** En $\mathbb{R}^4$ se consideran las variedades afines

$$L_1 = \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 : \; x_1 + x_2 = 4, \; x_3 + x_4 = a\}$$
$$L_2 = (3,2,2,-1) + \langle(1,-2,2,1)\rangle$$

Hallar $a$ para que $L_1$ y $L_2$ tengan intersección no vacía.

**E.10** Estudiar la posición relativa de los planos

$$\pi_1 \equiv -3x + 3y - 4z = 6, \qquad \pi_2 \equiv 4x - ky + 8z = 5$$

en función de los valores del parámetro $k$.

**E.11** Dadas las rectas

$$r \equiv \frac{x-2}{2} = \frac{y-1}{2} = z, \qquad s \equiv \frac{x-a}{a} = y - 1 = z + 1$$

Se pide:

(a) Calcula los valores de $a$ que hacen que las rectas anteriores se corten en un punto $P$ a determinar.

(b) Calcula la distancia del punto $P$ a la recta de ecuaciones

$$\begin{cases} x = 2 \\ x + y - 2z = 1 \end{cases}$$

(c) Calcula la perpendicular por $P$ al plano que pasa por los puntos $(0,1,0)$, $(1,1,0)$ y $(0,0,1)$.

**E.12** Hallar las ecuaciones de la transformación afín que asocia a cada punto de $\mathbb{R}^2$ su proyección ortogonal sobre la recta $y = 2x$.

**E.13** Calcula las ecuaciones de la proyección ortogonal sobre el plano de ecuación $x - 2y + z = 3$.

**E.14** Una *homotecia* de razón $k \neq 0$ es una transformación afín cuya matriz de la aplicación lineal asociada es $kI_n$. Calcula las ecuaciones de una homotecia de razón 2 que transforma el punto $(1,2,1)$ en el punto $(1,3,1)$. ¿Puedes describir geométricamente qué hace esta aplicación?

**E.15** Encontrar las ecuaciones de una aplicación afín que transforma los puntos $(0,0,0)$, $(1,0,0)$, $(1,1,0)$ y $(1,1,1)$ en los puntos $(1,1,1)$, $(1,2,3)$, $(1,2,4)$ y $(0,0,0)$, respectivamente.

## Ejercicios teóricos

**E.16** Dados $n+1$ puntos $P_0, P_1, \dots, P_n$, probar que los vectores $\overrightarrow{P_0P_1}, \overrightarrow{P_1P_2}, \dots, \overrightarrow{P_0P_n}$ son linealmente dependientes si y solo si, para cualquier $i = 1, \dots, n$ los $n$ vectores $\overrightarrow{P_iP_0}$, $\overrightarrow{P_iP_1}$, $\dots$, $\overrightarrow{P_iP_n}$ también lo son.

**E.17** En $\mathbb{R}^2$ considera un punto $P(x_0, y_0)$ y una recta $r \equiv ax + by + c = 0$. Deducir la fórmula de la distancia del punto $P$ a la recta $r$.

**E.18** En $\mathbb{R}^2$ se consideran las rectas $r \equiv ax+by+c = 0$ y $s \equiv ax+by+d = 0$, con $c \neq d$. Probar que las rectas son paralelas y encontrar una expresión que proporcione la distancia entre ellas.

## Ejercicios adicionales

**E.19** Probar que una aplicación lineal es ortogonal si y solo si conserva las distancias.

**E.20** Identificar la aplicación afín de matriz

$$\begin{pmatrix} 1 & 0 & 0 \\ 3 & -\frac{3}{5} & -\frac{4}{5} \\ 1 & -\frac{4}{5} & \frac{3}{5} \end{pmatrix}$$

**E.21** Construir las matrices de los siguentes movimientos:

(a) Una traslación de vector $(2,3)$.

(b) Un giro de centro el punto $(1, 3)$ y ángulo $\frac{\pi}{4}$.

(c) Una simetría de eje la recta $x - y = 2$.

# A Conceptos generales

En este apéndice tratamos de incluir una serie de nociones básicas que se suponen conocidas por el lector y que incluimos aquí a modo de repaso. En esencia recordaremos brevemente algunos conceptos de teoría de conjuntos, funciones y estructuras algebraicas, y veremos con más detalle el Principio de Inducción que es usado en diversas demostraciones a lo largo del texto.

## A 1 TEORÍA DE CONJUNTOS

Las nociones que siguen a continuación serán presentadas de forma "intuitiva", dado que una profundización en algunos de los conceptos que vamos a mostrar queda fuera del alcance de este texto.

El lector estará familiarizado con el símbolo lógico de *implicación*, $\Rightarrow$, que relaciona una afirmación con su consecuencia. La proposición $p \Rightarrow q$, que se lee **si** $p$ **entonces** $q$, significa que si la afirmación $p$ es cierta, entonces la afirmación $q$ también lo es. Por ejemplo,

**si** *llueve* **entonces** *la calles se mojan*

En este caso se dice que $p$ es una *condición suficiente* para $q$, mientras que $q$ se dice *condición necesaria* para $p$. Es importante señalar que si la afirmación $p$ no es cierta, no se puede deducir nada sobre la veracidad de $q$ (si no llueve, las calles podrían estar o no mojadas —por ejemplo, podrían haber regado—). Pero si $q$ es falsa, entonces $p$ no puede ser cierta (si las calles no están mojadas es porque no ha llovido).

Debemos recordar la diferencia entre la implicación simple y la *doble implicación* $p \Leftrightarrow q$, que se lee $p$ **si y solo si** $q$, y que significa

$$p \Rightarrow q \quad \text{y} \quad q \Rightarrow p$$

simultáneamente. En este caso se dice que $p$ es una *condición necesaria y suficiente* para $q$, o que $p$ y $q$ son *equivalentes*.

El lector también debe conocer el concepto de *conjunto* y de *elemento*, así como la relación de *pertenencia* a un conjunto, que denotamos por $x \in A$, y que significa que el elemento $x$ está en el conjunto $A$. La no pertenencia se

denota por $x \notin A$. Recordamos también la notación habitual para un conjunto formado por una serie de elementos:

$$A = \{a, b, c, d\} \tag{A.1}$$

Una colección de elementos de un conjunto se dice que forma un *subconjunto*, y en tal caso usamos la notación de *inclusión*, $B \subset A$, para decir que todo elemento del conjunto $B$ está en $A$. La inclusión *estricta* se denota por $B \subsetneq A$ y significa que todo elemento de $B$ está en $A$, pero hay elementos de $A$ que no están en $B$. Obsérvese que si $A$ es el conjunto dado en (A.1) entonces podemos escribir

$$a \in A, \quad b \in A, \quad \{b\} \subset A, \quad \{a, c\} \subset A$$

pero no $\{b, c, d\} \in A$. Asimismo, el lector podrá entender de inmediato la siguiente afirmación:

$$A \subset B \text{ y } B \subset A \Leftrightarrow A = B$$

Recordamos también la notación para el conjunto vacío $\emptyset$, y que se verifica que $\emptyset \subset A$ para cualquier conjunto $A$.

El conjunto formado por todos los subconjuntos de un conjunto $A$ (incluidos el conjunto vacío y el propio conjunto) se denomina *partes* de $A$ y se denota por $\mathcal{P}(A)$.

Usaremos a lo largo de todo el texto los *cuantificadores* $\forall$ y $\exists$, así como la negación de este último $\nexists$, y cuyos significados se muestran en los siguientes ejemplos:

$$p(x),\ \forall x \in A \qquad \text{se lee, } p(x) \textbf{ para todo } x \text{ de } A$$

y significa que la propiedad $p(x)$ se tiene para cualquier elemento $x$ del conjunto $A$.

$$\exists x \in A:\ p(x) \qquad \text{se lee, } \textbf{existe } x \text{ en } A \textbf{ tal que } p(x)$$

quiere decir que existe al menos un elemento $x$ del conjunto $A$ para el que la propiedad $p(x)$ se tiene.

Nótese que:

$$\nexists x \in A:\ p(x) \Leftrightarrow p(x) \text{ es falso } \forall x \in A$$

es decir, no hay ningún elemento de $A$ para el que la propiedad $p(x)$ sea cierta.

Es importante no confundir lo anterior con la negación de $p(x)$, $\forall x \in A$. Tal negación es equivalente a que $\exists x \in A$ tal que $p(x)$ se cumple.

Las operaciones entre conjuntos que usaremos son la *unión*, $A \cup B$, que corresponde el conjunto de elementos formados por todos los del conjunto $A$ y todos los del conjunto $B$, y la *intersección*, $A \cap B$, que se forma con los elementos que está a la vez en el conjunto $A$ y en el $B$. Se usarán también las uniones e

intersecciones genéricas

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$
$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

Por otro lado, la *diferencia* de conjuntos, $A \backslash B$, está formada por los elementos de $A$ que no están en $B$. Lógicamente se tiene que

$$A \backslash B \neq \emptyset \Rightarrow B \subset A$$

El *producto cartesiano* de dos conjuntos $A$ y $B$, que se denota por $A \times B$, se define como el conjunto de *pares ordenados* de la forma $(x, y)$ con $x \in A$ e $y \in B$. La definición se puede extender a un número mayor de conjuntos.

Una *relación* $R$ sobre un conjunto $A$ es un subconjunto del producto cartesiano $A \times A$. Si $x, y \in A$ son tales que el par $(x, y) \in R$, se representará por $x\,R\,y$, y se dirá que $x$ está relacionado con $y$ por la relación $R$.

Dada una relación $R$ en un conjunto $A$, se dice que ésta es

- *reflexiva*, si $x\,R\,x$, $\forall x \in A$
- *simétrica*, si $x\,R\,y \Rightarrow y\,R\,x$, $\forall x, y \in A$
- *antisimétrica* si $x\,R\,y$ y $y\,R\,x \Rightarrow x = y$, $\forall x, y \in A$
- *transitiva*, si $x\,R\,y$ y $y\,R\,z \Rightarrow x\,R\,z$, $\forall x, y, z \in A$

Una relación que sea reflexiva, antisimétrica y transitiva se dice *relación de orden*, o simplemente *orden*. Por ejemplo, la relación de inclusión entre conjuntos ($\subset$) es un orden en el conjunto $\mathcal{P}(A)$. La relación de desigualdad entre números ($\leq$), bien conocida por el lector, también es una relación de orden. Cuando existe una relación de orden en un conjunto se dice que éste es *ordenado*.

Una relación $R$ sobre un conjunto $A$ se dice de *equivalencia* si es reflexiva, simétrica y transitiva. Si $x\,R\,y$ con $R$ una relación de equivalencia, se dice que $x$ es equivalente a $y$ y se suele usar la notación $x \sim y$. Dado $x \in A$, el conjunto de elementos de $A$ que son equivalentes a $x$ se denomina *clase de equivalencia* de $x$. El conjunto formado por todas las clases de equivalencia se denomina *conjunto cociente* y se denota por $A/\sim$.

## A 2 FUNCIONES

Otro elemento fundamental con el que el lector probablemente estará familizarizado es el concepto de aplicación entre conjuntos.

Dados $S$ y $T$ dos conjuntos, toda correspondencia que asocia a cada elemento de $S$ un y solo un elemento de $T$ se denomina *aplicación* de $S$ en $T$. A $S$ se le llama *espacio de partida* o *espacio origen* y a $T$ *espacio de llegada* o *espacio imagen*.

La notación habitual para una aplicación es la siguiente:

$$\begin{array}{rccl} f: & S & \longrightarrow & T \\ & s & \longrightarrow & f(s) = t \equiv \text{imagen por } f \text{ de } s \end{array}$$

Cuando $S$ y $T$ son conjuntos numéricos se habla habitualmente de *función*.

Asociada a una función existen una serie de conjuntos y propiedades importantes; en concreto, si $f : S \longrightarrow T$ una aplicación, entonces:

(i) Se define el *dominio* de $f$, como el conjunto

$$\mathrm{Dom}(f) = \{s \in S : \exists f(s)\}$$

(ii) Se denomina *imagen* de $f$, al conjunto

$$\mathrm{Im}(f) = \{t \in T : \exists s \in S \text{ tal que } f(s) = t\} = \{f(s) : s \in S\} = f(S)$$

(iii) Si $\mathrm{Im}(f) = T$, es decir, si $\forall t \in T$, $\exists s \in S$ con $f(s) = t$, se dice que $f$ es *sobreyectiva* o *sobre*.

(iv) $f$ es *inyectiva* si

$$f(s) = f(s') \Longrightarrow s = s'$$

(v) $f$ se dirá *biyectiva* si es sobre e inyectiva.

Además existe una operación fundamental con las aplicaciones, la composición, que se define del siguiente modo: si $f : S \to T$ y $g : T \to U$ son tales que $\mathrm{Im}(f) \subset \mathrm{Dom}(g)$ se define la *composición* de $f$ con $g$, que se notará $(g \circ f)$, por:

$$\begin{array}{rccc} (g \circ f): & S & \longrightarrow & U \\ & s & \longrightarrow & g(f(s)) \end{array}$$

Es fácil demostrar que la composición de aplicaciones inyectivas es inyectiva, la composición de aplicaciones sobreyectivas es sobreyectiva, y obviamente, la composición de aplicaciones biyectivas es biyectiva.

Finalmente, dada una aplicación $f : S \to T$, se denomina *inversa* de $f$ a toda función $g : T \to S$ tal que

$$(g \circ f)(s) = s, \quad (f \circ g)(t) = t$$

Se notará $g = f^{-1}$. Es importante observar que la inversa no siempre existe. En caso de que exista se dirá que $f$ es *invertible*.

A 3

## ESTRUCTURAS ALGEBRAICAS

Se denomina *operación* (*interna*) sobre un conjunto $A$ a una aplicación

$$\star : A \times A \longrightarrow A$$

que a cada par de elementos $(x, y) \in A \times A$ le asocia un elemento $x \star y \in A$. Es decir, una operación consiste esencialmente en "hacer algo" con un par de elementos de un conjunto para "producir" un nuevo elemento del conjunto.

Se dice que una operación $\star$ es:

- *conmutativa*, si $x \star y = y \star x$, $\forall x, y \in A$
- *asociativa*, si $x \star (y \star z) = (x \star y) \star z$, $\forall x, y, z \in A$

Decimos que $e$ es un *elemento neutro* para la operación $\star$, si $x \star e = e \star x = x$, $\forall x \in A$.

Si la operación $\star$ posee elemento neutro $e$, dado $x \in A$ se dice que $x' \in A$ es su *simétrico* (también llamado *opuesto* o *inverso*) si $x \star x' = x' \star x = e$.

Cuando en un conjunto tenemos dos operaciones $+$ y $*$, se dice que $*$ es *distributiva por la izquierda* respecto de $+$ si

$$x * (y + z) = (x + y) * (x + z), \quad \forall x, y, z \in A$$

o *distributiva por la derecha* respecto de $+$ si

$$(y + z) * x = (y + x) * (z + x), \quad \forall x, y, z \in A$$

Obviamente, si $*$ es conmutativa y distributiva respecto de $+$ por alguno de los lados, lo será por el otro.

Una estructura algebraica debe entenderse como una "etiqueta" con la que se trata de identificar las propiedades que ciertas operaciones poseen en un determinado conjunto. El principal interés en trabajar con estructuras algebraicas estriba en el hecho de que las propiedades que se pueden deducir a partir de la estructura son independientes de la naturaleza de los elementos del conjunto en cuestión.

Si $A$ es un conjunto y $+$ es una operación asociativa y que posee elemento neutro y elemento simétrico, se dice que el par $(A, +)$ es un *grupo*. Si además la operación es conmutativa, se dice que es un *grupo conmutativo*.

Si en $A$ tenemos dos operaciones, $+$ y $*$, tales que $(A, +)$ es un grupo conmutativo, la operación $*$ es asociativa y distributiva respecto de $+$ tanto por la izquierda como por la derecha se dice que la terna $(A, +, *)$ es un *anillo*. El anillo es conmutativo si la operación $*$ lo es, y se dirá que es un *anillo unitario* si la operación $*$ posee elemento neutro.

Por ejemplo, el conjunto de números enteros $\mathbb{Z}$ es un anillo conmutativo y unitario.

Por último, un conjunto $A$ dotado de dos operaciones $+$ y $\cdot$ tales que $(A,+,\cdot)$ es un anillo y $(A^*,\cdot)$ es un grupo, donde $A^* = A\backslash\{0\}$, siendo 0 el elemento neutro para la operación $+$, se dice *cuerpo*.

Los ejemplos típicos de cuerpos que el lector probablemente conoce son $\mathbb{Q}$, el conjunto de números racionales, y $\mathbb{R}$, el conjunto de números reales.

A 4

## PRINCIPIO DE INDUCCIÓN

El Principio de Inducción es una técnica de demostración que se usa en varias ocasiones a lo largo del texto y que permite probar propiedades vinculadas de algún modo con el conjunto de números naturales $\mathbb{N}$.

Funciona del siguiente modo: si se pretende demostrar que una cierta propiedad $p(n)$ es cierta para todo número natural $n \in \mathbb{N}$, entonces es necesario probar:

(I) $p(1)$ es cierto

(II) $p(n) \Rightarrow p(n+1)$

Podemos tratar de entender el mecanismo si pensamos, por ejemplo, en subir una escalera. Para recorrer toda una escalera solo necesitamos realizar dos acciones: la primera y más evidente es llegar al primer escalón (lo que identificamos con probar que la propiedad es cierta en el caso $n = 1$, esto es, que $p(1)$ se cumple); la segunda consiste en, estando en un determinado escalón, saber llegar al escalón inmediatamente superior, es decir, si la propiedad en el caso $n$ es cierta, hemos de poder probar que también es cierta en el caso $n+1$ (lo que denotamos con $p(n) \Rightarrow p(n+1)$). Cómo hemos llegado al nivel $n$ es irrelevante aquí; lo que realmente importa es ver cómo llegar del nivel $n$ al nivel $n+1$, es decir, cómo se sube de un escalón al siguiente.

Debe ser ahora evidente que si somos capaces de llegar al primer escalón, y sabemos cómo pasar de un escalón al siguiente, entonces podemos subir completamente la escalera, no importa cuán larga sea. Veámoslo con un ejemplo.

**Ejemplo A.1**

(I) Probar que $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Es fácil comprobar que la fórmula funciona para algunos valores de $n$:

$$n = 1 \longrightarrow 1 = \frac{1\cdot 2}{2}$$

$$n = 2 \longrightarrow 1 + 2 = \frac{2 \cdot 3}{2}$$
$$n = 3 \longrightarrow 1 + 2 + 3 = \frac{3 \cdot 4}{2}$$

sin embargo es muy importante tener claro que esto no constituye una demostración del resultado (¿cómo podemos garantizar que la fórmula es válida para cualquier otro valor de $n$?). Vamos a probar la fórmula por inducción. El primer paso ya está hecho (la fórmula es cierta para $n = 1$, como acabamos de comprobar); veamos el segundo: hay que probar que si la fórmula es cierta en el caso $n$, entonces también lo es en el caso $n + 1$.

Escribamos cómo es la fórmula en el caso $n + 1$:

$$\sum_{i=1}^{n+1} i = \frac{(n+1)(n+2)}{2} \tag{A.2}$$

Esta expresión es la que queremos demostrar, por lo que comenzaremos desde un miembro de la igualdad y trataremos de llegar al otro. Aquí es importante señalar que estamos suponiendo que la fórmula es cierta en el caso $n$, es decir, que

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \tag{A.3}$$

se tiene. Esto puede ser confuso para el lector, pues parece que asumimos lo que queremos probar, pero no debemos olvidar que lo que debemos hacer es demostrar los pasos que marca el principio de inducción. Por tanto para probar (A.2):

$$\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}$$

Nótese que en el primer paso hemos separado la suma hasta $n + 1$ en la suma hasta $n$ más el término $n + 1$ y en el segundo paso usamos (A.3), a la que usualmente nos referiremos como *hipótesis de inducción*.

(II) Probar la fórmula del binomio de Newton:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Del mismo modo que antes, veamos en primer lugar que la fórmula es cierta para $n = 1$: en efecto,

$$\binom{1}{0} a^0 b^1 + \binom{1}{1} a^1 b^0 = a + b$$

Probemos ahora que si la fórmula es cierta en el caso $n-1$, entonces también es cierta en el caso $n$. Nótese que a diferencia del caso anterior en el que queríamos probar el caso $n+1$ dando por supuesto el caso $n$, ahora queremos probar el caso $n$ suponiendo el caso $n-1$. El lector debe entender que ambas situaciones demuestran el mismo hecho.

Así pues, para probar el caso $n$, consideramos $(a+b)^n$ y lo escribimos como $(a+b)^{n-1}(a+b)$, de modo que, usandola hipótesis de inducción:

$$\begin{aligned}(a+b)^{n-1}(a+b) &= (a+b)\sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-1-k} \\ &= a\sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-1-k} + b\sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-1-k} \\ &= \sum_{k=0}^{n-1}\binom{n-1}{k}a^{k+1} b^{n-1-k} + \sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-k}\end{aligned}$$

El primer sumando de esta última expresión se puede escribir del siguiente modo:

$$\sum_{k=0}^{n-1}\binom{n-1}{k}a^{k+1}b^{n-1-k} = \sum_{k=1}^{n}\binom{n-1}{k-1}a^k b^{n-k}$$

de modo que

$$\begin{aligned}(a+b)^n &= \sum_{k=1}^{n}\binom{n-1}{k-1}a^k b^{n-k} + \sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-k} \\ &= \binom{n-1}{0}b^n + \sum_{k=1}^{n-1}\left[\binom{n-1}{k-1}+\binom{n-1}{k}\right]a^k b^{n-k} + \binom{n-1}{n-1}a^n\end{aligned}$$

Nótese que la última línea se ha obtenido separando el término $k=0$ del segundo sumando, el término $k=n$ del primero y agrupando el resto de los sumatorios. Finalmente, basta ver[1] que

$$\binom{n-1}{k-1}+\binom{n-1}{k}=\binom{n}{k}$$

y que

$$\binom{n-1}{0}=\binom{n}{0}=1 \qquad \binom{n-1}{n-1}=\binom{n}{n}=1$$

para obtener el resultado deseado.

# B Introducción a Python

Este tema no pretende ser un manual de uso de Python sino una breve introducción al mismo que proporcione al lector un punto de partida a partir del cual pueda usar el lenguaje para abordar una buena parte de los cálculos, en algunos momentos tediosos, que son precisos realizar a lo largo de este texto.

Python es un lenguaje de programación creado por Guido Van Rossum a finales de los ochenta. Su nombre deriva de la afición de su creador al grupo de humor inglés *Monty Python*. Python es un lenguaje de alto nivel, interpretado, interactivo y de propósito general. Un lenguaje se dice de *alto nivel* cuando su léxico y estructura está más próximo al lenguaje humano que al lenguaje que entiende el computador. El lenguaje es *interpretado* cuando no necesita de un proceso de compilación.[1] En estos casos, el intérprete lee las líneas de código y las ejecuta una a una. Es *interactivo* porque proporciona la posibilidad de ejecutar código directamente sobre el intérprete sin necesidad de escribir un programa, capacidad ésta, que nos será de especial utilidad a lo largo del texto, y es un lenguaje de *propósito general* puesto que es lo suficientemente versátil como para programar cualquier tarea que pueda realizarse con un ordenador. Como carta de presentación podemos decir que Python es uno de los tres lenguajes oficiales empleados en Google.

Los lectores con conocimiento de algún lenguaje de programación encontrarán en Python un lenguaje sencillo, versátil y que proporciona código fácilmente legible. Para aquéllos que no están familiarizados con la programación, Python supone un primer contacto agradable, pues los programas pueden ser comprobados y depurados con facilidad, permitiendo al usuario concentrarse más en el problema a resolver que en los aspectos concretos de la programación.

Python es software de código abierto que está disponible en múltiples plataformas (Linux, Unix, Windows, Mac OS, etc.), de manera que el mismo código funciona en diferentes sistemas,[2] aunque para ello es preciso disponer del intérprete. En este tema veremos algunos aspectos generales relacionados con la instalación y uso del intérprete, así como las características básicas del lenguaje y los paquetes esenciales que nos serán de utilidad a lo largo del texto.

[1] Los lenguajes compilados precisan de un proceso de *compilación* (y posterior enlazado) que transforman el código escrito por el programador (el llamado código *fuente*) en un código ejecutable (binario).

[2] Con la excepción de las extensiones que son específicas de cada sistema operativo.

B 1

## INSTALACIÓN DE PYTHON

Python puede ser obtenido desde la página web del lenguaje.[3] El proceso de instalación dependerá del sistema empleado y no suele presentar complicaciones. Tanto en Windows como en Mac, bastará con bajar el instalador y ejecutarlo.[4] En la mayoría de distribuciones Linux, Python viene instalado por defecto o puede ser instalado fácilmente con la herramienta de gestión de paquetes.

Una de las principales virtudes de Python está en la gran cantidad de extensiones del lenguaje que posee. Estas extensiones suponen complementos adecuados para realizar de forma rápida y sencilla ciertas operaciones que no forman parte del núcleo principal del lenguaje. Así, en este texto trataremos dos extensiones que nos serán de gran utilidad: NumPy y SymPy. La primera de ellas está diseñada para trabajar con cálculo vectorial avanzado, mientras que la segunda incorpora a Python elementos de cálculo simbólico que nos permitirán realizar operaciones matemáticas con expresiones simbólicas y cálculos exactos. Estas dos extensiones no vienen con la instalación inicial de Python y deben ser instaladas posteriormente. Pueden ser descargadas desde `numpy.org` y `sympy.org`, respectivamente.[5]

### 2 1 1 Primeros pasos

Una vez instalado el intérprete de Python podemos trabajar con el lenguaje de dos formas distintas: a través de la consola o mediante la ejecución de *scripts* de órdenes. El primer método es bastante útil cuando queremos realizar operaciones inmediatas, y podemos compararlo con el uso de una calculadora avanzada. El uso de *scripts* de órdenes corresponde a la escritura de código Python que es posteriormente ejecutado a través del intérprete.

Hay diversas formas de iniciar el intérprete Python: en Windows lo haremos directamente abriendo la consola Python o el programa IDLE Python en el menú de programas, mientras que en Linux o Mac OS abriremos la típica *terminal* y ejecutaremos la orden `python`. El resultado será algo por el estilo:

```
Python 2.6.5 (r265:79063, Apr 16 2010, 13:09:56)
[GCC 4.4.3] on linux2
Type "help", "copyright", "credits" or "license" for more
    information.
>>>
```

que nos informa de la versión que tenemos instalada y nos señala el *prompt* `>>>` del sistema, el cual indica la situación del terminal a la espera de órdenes.

[3] `www.python.org/download`

[4] En las últimas versiones de los sistemas Mac ya viene instalado por defecto.

[5] La descarga e instalación en una distribución Ubuntu se hace sencillamente con la orden `sudo apt-get install python-numpy python-sympy`

Podemos salir con la orden `exit()` o pulsando las teclas `Ctrl+D` (en Linux y Mac) ó `Ctrl+Z` y ↲[6] (en Windows).

Una vez dentro del intérprete podemos ejecutar órdenes del sistema, por ejemplo

```
>>> print "Hola Mundo"
Hola Mundo
>>>
```

Obviamente la orden `print` imprime la *cadena de texto* o *string* `Hola Mundo` que va encerrada entre comillas para indicar precisamente que se trata de un *string*. Una vez ejecutada la orden, el sistema vuelve a mostrar el *prompt*.

La otra alternativa a la ejecución de órdenes con Python es la creación de un *script* o *guión*. Se trata de un archivo de texto en el que listamos las órdenes Python que pretendemos ejecutar. Para la edición del archivo nos vale cualquier editor de texto sin formato. Escribiendo el comando

```
print "Hola Mundo"
```

en un archivo,[7] lo salvamos con un nombre cualquiera, por ejemplo `hola.py`. Desde Windows podemos ejecutar el *script* haciendo doble click sobre el archivo en cuestión, lo que ocurre es que se mostrará el resultado en una ventana de comandos rápidamente y ésta se cerrará, lo cual no nos permitirá ver nada. Esto puede se arreglado añadiendo al fichero la orden `raw_input()`, que hará que el programa se quede a la espera de una entrada por teclado para cerrarse después.

En los sistemas Linux o Mac, podemos ejecutar el código sencillamente escribiendo en una consola la orden `python hola.py` (obviamente situándonos correctamente en el *path*[8]). También es posible hacer ejecutable el código Python escribiendo en la primera línea del archivo[9]

```
#!/usr/bin/env python
```

y dando permisos de ejecución al archivo con la orden `chmod a+x hola.py` desde una consola. En tal caso podemos ejecutar con un doble click sobre el archivo o escribiendo `./hola.py` en una consola.

Se pueden utilizar codificaciones diferentes de la ASCII[10] en los *scripts* de Python añadiendo, justo detrás del *shebang* la línea

---

[6] El símbolo ↲ denotará la tecla `Enter` o retorno de carro.

[7] Para diferenciar la escritura de órdenes en el intérprete de los comandos que introduciremos en los archivos ilustraremos los últimos con un marco doble.

[8] El *path* o *ruta* de un archivo nos indica la localización de éste dentro del sistema de archivos.

[9] Esto es lo que se conoce como el *shebang*, y es el método estándar para poder ejecutar un programa interpretado como si fuera un binario.

[10] Es decir, codificaciones que admiten caracteres acentuados.

```
# -*- coding: codificación -*-
```

donde *codificación* se refiere al código de caracteres que empleemos (típicamente `utf-8` en Linux o Mac). El empleo de caracteres no ASCII en un *script* sin esta línea produce errores.

### 2 1 2 La consola iPython

En lugar del intérprete Python habitual existe una consola interactiva denominada iPython con una serie de características muy interesantes que facilitan el trabajo con el intérprete.[11]

La principal virtud de esta consola es la presencia del *autocompletado*, característica que se activa al pulsar la tecla de tabulación y que nos permite que al teclear las primeras letras de una orden aparezcan todas las órdenes disponibles que comienzan de esa forma. También existe un operador `?` que puesto al final de una orden nos muestra una breve ayuda acerca de dicha orden. Además, esta consola pone a nuestra disposición comandos del entorno (`cd`, `ls`, etc.) que nos permiten movernos por el árbol de directorio desde dentro de la consola, y el comando `run` con el que podemos ejecutar desde la consola un *script* de órdenes.

## B 2 ASPECTOS BÁSICOS DEL LENGUAJE

### 2 2 1 Variables y tipos de datos

En un lenguaje de programación, una *variable* es un símbolo que representa algún tipo de dato, ya sea un número entero, un número real, una cadena de caracteres, etc. Python es lo que se conoce como un lenguaje *dinámicamente tipado*, es decir, las variables pueden cambiar de tipo en distintos momentos sin necesidad de ser previamente declaradas. Las variables son identificadas con un nombre, mediante el cual podemos acceder al dato que almacenan o modificarlo. Los nombres de variables deben obligatoriamente comenzar por una letra y hay distinción entre mayúsculas y minúsculas.

Los tipos de datos que usaremos en este texto son los números enteros (`int`), reales en doble precisión (`float`), complejos (`complex`), cadenas de caracteres (`str`) y listas (`list`).

Las variables se definen con el operador de asignación `=`; por ejemplo

```
>>> a=2
>>> b=5.
>>> c=3+1j
```

[11] Se puede descargar desde `ipython.scipy.org`.

define las variables `a`, `b` y `c` como un entero de valor 2, un real de valor 5 y un número complejo de valor $3 + i$, respectivamente. Obsérvese la necesidad de poner un punto para definir el valor como real y no como entero y el uso de `j` en lugar de $i$ en el número complejo, así como la obligación de anteponer un número. También se puede declarar un número complejo del siguiente modo:

```
>>> complex(3,2)
(3+2j)
```

Podemos recuperar el tipo de dato de cada variable con la orden `type`,

```
>>> type(a)
<type 'int'>
>>> type(b)
<type 'float'>
>>> type(c)
<type 'complex'>
```

Como vemos, Python asigna el tipo a cada variable en función de su definición. Es importante resaltar la diferencia entre los tipos numéricos, pues si no somos cuidadosos podemos caer en el siguiente error:

```
>>> a=5; b=2
>>> a+b
7
>>> a/b
2
```

Obsérvese que hemos definido las variables `a` y `b` en la misma línea, separadas por `;`. Claramente `a+b` calcula la suma de los valores de las variables, sin embargo `a/b` parece que no calcula correctamente la división. En realidad la respuesta es correcta dado que ambas variables son enteros, y por tanto se realiza la división entre enteros, que corresponde a la parte entera de la división. Si lo que esperamos es obtener la división real debemos escribir al menos uno de los números en forma real, lo que se hace con el comando `float`

```
>>> a/float(b)
2.5
```

Cuando Python opera con números de distinto tipo, realiza la operación transformando todos los números involucrados al mismo tipo según una jerarquía establecida que va de enteros a reales y luego a complejos. Obsérvese el siguiente ejemplo:

```
>>> a=3.
>>> b=2+3j
>>> c=a+b
>>> c
(5+3j)
```

```
>>> type(c)
<type 'complex'>
```

Como vemos, la operación de suma entre un real y un complejo se realiza correctamente como suma de números complejos y el resultado corresponde a un número de ese tipo.

Las *cadenas* no son más que texto encerrado entre comillas:

```
>>> a="Hola"
>>> b='mundo'
>>> a
'Hola'
>>> b
'mundo'
>>> type(a)
<type 'str'>
```

en las que se puede comprobar que da igual definirlas con comillas simples o dobles.

Por último,[12] las *listas* son colecciones de variables de cualquier tipo (inclusive listas). Obsérvese el siguiente ejemplo:

```
>>> a=[1,2.,3+1j,"hola"]
>>> type(a)
<type 'list'>
>>> a
[1, 2.0, (3+1j), 'hola']
>>> a[1]
2.0
>>> type(a[1])
<type 'float'>
>>> a[0]
1
```

Hemos definido una lista encerrando sus elementos (de tipos diversos) entre corchetes y separándolos por comas. Podemos acceder a cada uno de los elementos de la lista escribiendo el nombre de la lista y el índice del elemento entre corchetes, teniendo en cuenta que el primer elemento tiene índice 0. Si algún elemento de la lista es otra lista, podemos acceder a los elementos de esta última usando el corchete dos veces, como en el siguiente ejemplo:

```
>>> a=[1,2.,3+1j,"hola"]
>>> milista=[1,2,a]
>>> milista
[1, 2, [1, 2.0, (3+1j), 'hola']]
>>> milista[2][3]
```

[12] Hay más tipos de datos en Python, pero de momento solo usaremos los aquí descritos.

```
'hola'
```

### 2.2.2 Operadores

Las operaciones numéricas más comunes se realizan en Python con los operadores `+` (suma), `-` (resta), `*` (multiplicación), `/` (división) y `**` (potenciación). También existe el operador `//` (división entera), que da la parte entera de la división entre dos reales y el operador `%` (módulo), que proporciona el resto de la división entre dos números.

Los operadores suma y multiplicación pueden aplicarse a cadenas y a listas, como vemos en el siguiente ejemplo:

```
>>> a="hola"; b=" amigo"
>>> a+b
'hola amigo'
>>> a*3
'holaholahola'
>>> a=[1,5,3]
>>> 3*a
[1, 5, 3, 1, 5, 3, 1, 5, 3]
>>> a+a
[1, 5, 3, 1, 5, 3]
```

Pero además, con las listas disponemos de operadores que nos permiten acceder no solo a elementos concretos de la lista, sino a trozos, lo que se denomina *slicing*. Veamos el siguiente ejemplo en el que se usa el comando `range`, cuyo funcionamiento es claro:

```
>>> a=range(10)
>>> a
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
>>> a[2:6]
[2, 3, 4, 5]
>>> a[:3]
[0, 1, 2]
>>> a[7:]
[7, 8, 9]
```

Nótese como el operador de *slicing* `inicio:fin` permite recuperar una parte de una lista entre los valores `inicio` y `fin`, sin contar este último. Más adelante veremos su utilidad en el manejo de submatrices.

## B.3 BUCLES Y CONDICIONALES

Las órdenes que vamos a ver en esta sección están orientadas a la escritura de *scripts* y a la realización de programas aunque también pueden usarse en el modo

interactivo. Los bucles permiten ejecutar fragmentos de código repetidamente para una serie de valores. Una característica esencial de Python es que la sintaxis del lenguaje impone obligatoriamente que escribamos con cierta claridad. Así, el fragmento de código a repetir en un bloque `for` está marcado por el sangrado de línea:

```
>>> for i in range(3):
...     print i
...
0
1
2
```

La sintaxis de la orden `for` es simple: la variable `i` recorre la lista generada por `range(3)`, finalizando con dos puntos (`:`) obligatoriamente. La siguiente línea, comenzada por `...` por el intérprete, debe ser sangrada, bien con el tabulador, bien con espacios (uno es suficiente, aunque lo habitual es cuatro). Al dejar la siguiente línea en blanco el intérprete entiende que hemos finalizado el bucle `for` y lo ejecuta. En un *script* volveríamos al sangrado inicial para indicar el fin del bucle.

Nótese que el bucle en Python corre a través de la lista y no de los índices de ésta, como se muestra en el siguiente ejemplo:

```
>>> a=['hola','mundo']
>>> for b in a:
...    print b
...
hola
mundo
```

La escritura de sentencias condicionales es similar a la de los bucles `for`, usando el sangrado de línea para determinar el bloque

```
>>> if 4%3 == 0:
...     print "4 es divisible por 3"
... else:
...     print "4 no es divisible por 3"
...
4 no es divisible por 3
```

La orden `if` evalúa la operación lógica "el resto de la división de 4 entre 3 es igual a cero", puesto que la respuesta es negativa, se ejecuta la segunda sentencia de impresión (`else`). Los *operadores relacionales* en Python son `==` (igual), `!=` (distinto), `>` (mayor que), `>=` (mayor o igual que), `<` (menor que) y `<=` (menor o igual que), y los *operadores lógicos* son `and`, `or` y `not`.

B 4

## MÓDULOS

Una de las principales características de Python es su extensa biblioteca que nos provee de funciones para realizar las tareas más diversas. Esta modularización del lenguaje permite además que los programas creados puedan ser reutilizados con facilidad. Para cargar un módulo usamos la orden `import`:

```
>>> import os
```

En este ejemplo hemos cargado el módulo `os` (*Operating System*) que nos provee de funciones del sistema operativo con las que podemos, por ejemplo, movernos a un directorio. El comando `dir(os)` nos proporciona un listado de los objetos, atributos y funciones del módulo. Para usar cualquier función del módulo es preciso anteponer su nombre. Por ejemplo,

```
>>> os.getcwd()
'/home/usuario'
```

nos proporciona una cadena de texto correspondiente al directorio en el que nos encontramos.[13] El comando `chdir` nos permite cambiar de directorio, tanto en forma absoluta como relativa, teniendo en cuenta que usamos la separación de directorios con la barra usual (/) en todos los sistemas:

```
>>> os.listdir('.')
['dir1','dir2','archivo']
>>> os.chdir('dir1')
>>> os.getcwd()
'/home/usuario/dir1'
```

Python está diseñado para ser usado a través de módulos, es decir, solo unas pocas funciones son cargadas con el núcleo principal, y para acceder a cálculos comunes, como por ejemplo la raíz cuadrada (`sqrt`), es necesario cargar el módulo apropiado.

Nótese que para usar las funciones del módulo cargado hemos de anteponer el nombre del módulo seguido de un punto. Podemos evitar el tener que hacer esto si cargamos los módulos del siguiente modo:

```
>>> from math import *
```

Con esta orden importamos todas las funciones del módulo `math` (básicamente las trigonométricas, logaritmo, exponencial, raíz cuadrada, número $\pi$, etc.), las cuales estarán accesibles sin necesidad de anteponer el nombre del módulo. Ahora, si queremos calcular $\sqrt{2}$, escribimos

[13] La respuesta en un sistema Windows podría ser algo similar a `'C:\\Python27'`.

```
>>> sqrt(2)
1.4142135623730951
```

Lógicamente, esta forma de cargar los módulos tiene ventajas evidentes en cuanto a la escritura de órdenes, pero tiene también sus inconvenientes. Por ejemplo, es posible que haya más de un módulo que use la misma función, como es el caso de la raíz cuadrada, que aparece tanto en el módulo `math` como en el módulo `cmath`. De manera que podemos encontrarnos situaciones como la siguiente:

```
>>> import math
>>> import cmath
>>> math.sqrt(-1)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ValueError: math domain error
>>> cmath.sqrt(-1)
1j
```

Como vemos, hemos cargado los módulos `math` y `cmath` y calculado la raíz cuadrada de $-1$ con la función `sqrt` que posee cada módulo. El resultado es bien distinto: la función raíz cuadrada del módulo `math` no permite el uso de números negativos, mientras que la función `sqrt` del módulo `cmath` sí. Si en lugar de cargar los módulos como en el último ejemplo los hubiésemos cargado así:

```
>>> from math import *
>>> from cmath import *
```

¿qué ocurrirá al hacer `sqrt(-1)`? Como el lector puede imaginar, la función `sqrt` del módulo `math` es sobreescrita por la del módulo `cmath`, por lo que solo la última es accesible.

Existe una tercera opción para acceder a las funciones de los módulos que no precisa importarlo al completo. Así,

```
>>> from math import sqrt
>>> from cmath import cos,sin
```

nos deja a nuestra disposición la función raíz cuadrada del módulo `math` y las funciones trigonométricas seno y coseno del módulo `cmath`. Es importante señalar que con estas funciones no tenemos acceso a ninguna otra función de los módulos que no hubiera sido previamente importada.

Esta última opción es de uso más frecuente en los *scripts*, debido a que con ella cargamos exclusivamente las funciones que vamos a necesitar y de esa forma mantenemos el programa con el mínimo necesario de recursos. En el uso de la consola interactiva es más frecuente cargar el módulo al completo, y es

aconsejable hacerlo sin el uso de `*`. De hecho, hay una posibilidad adicional que nos evita tener que escribir el nombre del módulo al completo, seguido del punto para usar una función. Si realizamos una importación del módulo como sigue.

```
>>> import math as m
```

entonces no es necesario escribir `math.` para acceder a la funciones sino

```
>>> m.cos(m.pi)
-1.0
```

### 2 4 1 Funciones definidas por el usuario

Las funciones son trozos de código que realizan una determinada tarea. Vienen definidas por la orden `def` y a continuación el nombre que las define. Siguiendo la sintaxis propia de Python, el código de la función está sangrado. La función finaliza con la orden `return`.

La principal característica de las funciones es que permiten pasarles argumentos de manera que la tarea que realizan cambia en función de dichos argumentos.

```
>>> def mifuncion(x,y):
...     t=x**2+y**2
...     return t
...
>>> mifuncion(2,3)
13
>>> mifuncion(4,-1)
17
```

Del mismo modo, como puede verse en el ejemplo, las funciones pueden devolver valores a través de la orden `return`.

Aunque las funciones pueden ser definidas dentro del intérprete para su uso, es más habitual almacenarlas en un fichero, bien para poder ser ejecutadas desde el mismo, o bien para ser importadas como si se tratara de un módulo. Por ejemplo, podemos definir una función matemática y guardarla en un archivo `tang.py`

```
from math import sin,cos,pi
def tangente(x):
    if cos(x)!=0:
        return sin(x)/cos(x)
    else:
        print "La tangente es infinita"
        return

x=tangente(pi)
print 'La tangente de pi es',x
```

Si ahora ejecutamos el archivo desde la consola:

```
$ python tang.py
La tangente de pi es -1.22460635382e-16
```

Aunque también podemos cargar la función tangente desde el intérprete como si se tratara de un módulo:[14]

```
>>> from tang import tangente
La tangente de pi es -1.22460635382e-16
>>> tangente(3)
-0.1425465430742778
```

Nótese cómo al cargar el módulo hacemos referencia al nombre del fichero, y éste es ejecutado. Posteriormente dispondremos de la función `tangente` para su uso.

### 2 4 2 Funciones, métodos y atributos

Aunque no es la intención de este texto introducir al lector en cuestiones relacionadas con la Programación Orientada a Objetos, es inevitable usar en algún momento algo de su terminología. De manera muy superficial podemos decir que los *objetos* equivalen a una especie de dato particular en Python, mientras que las funciones, los métodos y los atributos son esencialmente *operaciones* que pueden hacerse con estos objetos que tenemos a nuestra disposición. La diferencia entre unos y otros está en la forma en la que se accede a ellos. Ya hemos visto en ejemplos anteriores cómo trabajar con las funciones. A diferencia de éstas, los métodos y atributos no precisan importación, pues vienen definidos con el propio objeto. El siguiente ejemplo nos muestra la diferencia entre ambos:

```
>>> a=3+2j
>>> a.real
3.0
>>> a.conjugate
<built-in method conjugate of complex object at 0xb7738968>
>>> a.conjugate()
(3-2j)
```

Como vemos, el atributo `real` nos permite obtener la parte real del número complejo definido en el objeto `a`. Si queremos obtener su conjugado, vemos que `a.conjugate` nos informa que eso es un método, pero no nos proporciona el valor esperado. Para ello hemos de usar la notación `a.conjugate()`.

[14]Hay que tener la precaución de situar el fichero `tang.py` donde el intérprete pueda encontrarlo, por ejemplo en el directorio desde el que ejecutamos el intérprete.

## 2 4 3 El módulo NumPy

NumPy es un módulo fundamental para el cálculo científico con Python, principalmente porque está diseñado para obtener un buen nivel de rendimiento (cercano a lo que pueden proporcionarnos lenguajes compilados como FORTRAN o C) cuando manejamos un gran número de datos. Aunque en este texto no llevaremos a cabo cálculos en los que el rendimiento pueda ser un factor a tener en cuenta, cualquier aplicación avanzada en matemáticas requiere de un tratamiento eficiente en los cálculos, y en ese sentido NumPy es actualmente la opción más adecuada para ello.

El módulo NumPy incorpora un nuevo tipo de dato, el *array* o arreglo, que tiene una estructura similar a una lista, aunque es más restrictivo y computacionalmente mucho más eficiente. A diferencia de las listas, cuyos elementos pueden ser de cualquier tipo, un *array* de NumPy debe tener todos sus elementos del mismo tipo. Para definirlo, obviamente será necesario importar el módulo previamente,

```
>>> import numpy as np
>>> a=np.array([1,4,9])
>>> a
array([1, 4 ,9])
```

El acceso a los elementos del arreglo es idéntico al de las listas. Es muy importante tener presente el tipo de valores definidos en el arreglo, pues éste no cambia al modificar los elementos de tipo. Es decir, el arreglo definido antes tiene todos sus elementos enteros, como podemos comprobar con el atributo `dtype`

```
>>> a.dtype
dtype('int32')
```

La respuesta `int32` se refiere a un tipo de entero.[15] Si ahora pretendemos modificar los valores del arreglo, podemos llevarnos alguna sorpresa

```
>>> a[0]=3.6
>>> a
array([3, 4 ,9])
```

Como vemos, al ser el arreglo entero, la modificación del primer valor de éste toma solo su parte entera. Si queremos que el arreglo admita valores reales, hemos de modificar previamente su tipo:

```
>>> a=a.astype(float)
>>> a
array([ 3.,  4.,  9.])
```

[15] En NumPy existen varios tipos de enteros, reales y complejos, en función del número de cifras que queramos manejar. En todo este texto trabajaremos con los tipos que NumPy define por defecto.

```
>>> a[0]=3.6
>>> a
array([ 3.6,  4.,  9.])
```

Para evitarnos el tener que cambiar de tipo lo más sencillo es definir desde el principio el arreglo usando reales (bastará usarlo en una única entrada), es decir

```
>>> a=np.array([2.,5,7])
>>> a.dtype
dtype('float64')
```

### Matrices

Los arreglos también pueden ser multidimensionales; en particular, podemos trabajar con matrices como arreglos bidimensionales, definiéndoles del siguiente modo:

```
>>> M=np.array([[1.,3,5],[2,1,8]])
>>> M
array([[ 1.,  3.,  5.],
       [ 2.,  1.,  8.]])
```

y podemos acceder a sus elementos con un doble índice, tal y como haríamos con una matriz (recordando que los índices comienzan en 0):

```
>>> M[0,2]
5.0
```

Si bien trabajar con los arreglos multidimensionales es sencillo, NumPy incorpora un tipo especial de dato para trabajar con matrices, que será el que usemos en todo este texto, por estar especialmente diseñado para el álgebra lineal. Las matrices se pueden definir de varias formas, aquí mostramos algunos ejemplos:

```
>>> a=np.array([[1,2.,3],[2,5,2]])
>>> A=np.matrix(a)
>>> B=np.matrix([[1,0.,1],[2,0,-1]])
>>> C=np.matrix('1 3 6.; 0 1 1')
```

Obsérvese que la matriz $A$ se ha construido a partir de un arreglo bidimensional, la matriz $B$ se ha obtenido usando el mismo tipo de definición que los arreglos, mientras que $C$ se ha construido de forma más sencilla, como una especie de lista, separando los elementos de cada fila por espacios y los de cada columna por `;`. En el tema 2 tratamos las diversas operaciones y métodos existentes para las matrices.

### 2 4 4 El módulo SymPy

El cálculo numérico con un ordenador lleva aparejado ciertos inconvenientes debido a la imposibilidad de almacenar un numero infinito de cifras decimales. Así, el número $\frac{1}{3}$ no puede almacenarse más que a través de una aproximación numérica con un número finito de cifras decimales. Este hecho da lugar a los denominados *errores de redondeo* que tratamos en el tema 3. Los programas para el *cálculo simbólico* permiten que los ordenadores puedan trabajar, no con aproximaciones numéricas, sino con el número exacto, al igual que podemos hacer nosotros "a mano". El lector quizás se pregunte que, si esto es posible, por qué los ordenadores no trabajan siempre con cálculo simbólico. La respuesta es que el cálculo simbólico es mucho más lento que el cálculo numérico y no es posible trabajar con un número grande de datos.

SymPy es una librería de cálculo simbólico que no solo nos permitirá hacer cálculos de forma exacta, sino trabajar también con expresiones simbólicas, es decir, "con letras", como podemos hacer nosotros. Por ejemplo,

```
>>> import sympy as sp
>>> x,y=sp.symbols('x y')
>>> a=(x+y)**2
>>> sp.simplify(a)
2*x*y + x**2 + y**2
```

Hemos definido dos variables simbólicas $x$ e $y$ mediante la función `symbols`, de modo que la expresión $(x + y)^2$ ahora toma pleno sentido. Con la función `simplify` obtenemos dicha expresión simplificada. Este módulo también posee una forma de definir matrices simbólicas, aunque las formas cambian:

```
>>> A = sp.Matrix([[1,x], [y,1]])
>>> B=sp.Matrix(2,3,[1,2,x,0,1,y])
>>> B
[1, 2, x]
[0, 1, y]
```

y posibilita el cálculo exacto del siguiente modo:

```
>>> a=sp.Rational(1,2)
>>> b=sp.Rational(2,3)
>>> a+b
7/6
```

A lo largo de los diferentes temas iremos viendo algunas de las capacidades de este módulo.

# C Soluciones a los ejercicios de repaso

C 1

## NÚMEROS COMPLEJOS

**1.1** $4-7i$; $2+i$; $2$; $1+17i$; $2$; $-45+17i$;
$-\frac{13}{29}+\frac{11}{29}i$; $-\frac{23}{34}-\frac{41}{34}i$; $-\frac{1}{5}+\frac{2}{5}i$; $-2-5i$; $-\frac{3}{5}+i$; $4+i$

**1.2** $-i$; $3+2i$; $\pi+e+i$; $e-\pi$; $i$; $-1-\frac{3}{2}i$

**1.3** $2,\ \pi$; $1,\ -\frac{\pi}{2}$; $5\sqrt{2},\ -\frac{\pi}{4}$; $6,\ \frac{5\pi}{6}$; $2,\ -\frac{2\pi}{3}$; $2\sqrt{2},\ -\frac{3\pi}{4}$

**1.4**

$$2=2(\cos\pi+i\operatorname{sen}\pi);\quad -i=e^{-i\frac{\pi}{2}}=\cos(-\tfrac{\pi}{2})+i\operatorname{sen}(-\tfrac{\pi}{2})$$
$$5-5i=5\sqrt{2}e^{-i\frac{\pi}{4}}=5\sqrt{2}\left(\cos(-\tfrac{\pi}{4})+i\operatorname{sen}(-\tfrac{\pi}{4})\right)$$
$$-3\sqrt{3}+3i=6e^{\frac{5\pi}{6}}=6\left(\cos(\tfrac{5\pi}{6})+i\operatorname{sen}(\tfrac{5\pi}{6})\right)$$
$$-1-\sqrt{3}i=2e^{-\frac{2\pi}{3}}=2\left(\cos(-\tfrac{2\pi}{3})+i\operatorname{sen}(-\tfrac{2\pi}{3})\right)$$
$$-2-2i=2\sqrt{2}e^{-i\frac{3\pi}{4}}=2\sqrt{2}\left(\cos(-\tfrac{3\pi}{4})+i\operatorname{sen}(-\tfrac{3\pi}{4})\right)$$

**1.5** $-1$; $\frac{3\sqrt{2}}{2}+\frac{3\sqrt{2}}{2}i$; $-\sqrt{2}-\sqrt{2}i$; $\frac{\sqrt{3}}{4}+\frac{1}{4}i$; $-\frac{\sqrt{2}}{3}-\frac{\sqrt{2}}{3}i$; $-\frac{1}{2}+\frac{\sqrt{3}}{2}i$

**1.6** $i$; $16+16\sqrt{3}i$; $4096$; $-\frac{1}{2}+\frac{\sqrt{3}}{2}i$

**1.7**

$$\sqrt[4]{2},\ -\sqrt[4]{2},\ \sqrt[4]{2}i,\ -\sqrt[4]{2}i;\quad -\tfrac{\sqrt{3}}{4}+\tfrac{1}{4}i,\ \tfrac{\sqrt{3}}{4}+\tfrac{1}{4}i,\ -\tfrac{1}{2}i;$$
$$\tfrac{\sqrt{3}}{2}+\tfrac{1}{2}i,\ \tfrac{\sqrt{3}}{2}-\tfrac{1}{2}i,\ -\tfrac{\sqrt{3}}{2}+\tfrac{1}{2}i,\ -\tfrac{\sqrt{3}}{2}-\tfrac{1}{2}i,\ i,\ -i;$$

**1.8**

$$\sqrt{5}i,\ -\sqrt{5}i;\quad -\tfrac{1}{2}+\tfrac{\sqrt{3}}{2}i,\ -\tfrac{1}{2}-\tfrac{\sqrt{3}}{2}i;$$
$$0,\ -\tfrac{1}{2}+\tfrac{\sqrt{3}}{2}i,\ -\tfrac{1}{2}-\tfrac{\sqrt{3}}{2}i;\quad -1+2i,\ -1-2i;$$
$$3,\ \sqrt{5}i,\ -\sqrt{5}i;\quad -1,\ -\tfrac{1}{2}+\tfrac{\sqrt{11}}{2}i,\ -\tfrac{1}{2}-\tfrac{\sqrt{11}}{2}i;$$

C 2

## MATRICES Y DETERMINANTES

**2.1**

(a)

$$R_1 = \begin{pmatrix} 24500 & 19000 & 11200 \\ 14750 & 11500 & 7000 \\ 5125 & 4000 & 2450 \end{pmatrix} \qquad R_2 = \begin{pmatrix} 21250 & 22400 \\ 16125 & 17200 \end{pmatrix}$$

(b) $(R_2)_{22} = 17200$; $(R_1)_{33} = 2450$

(c) $(R_2)_{11}$

(d) $(R_1)_{11} + (R_1)_{22} + (R_1)_{33} = (R_2)_{11} + (R_2)_{22} = 38450$

**2.2** $\begin{pmatrix} 8 \end{pmatrix}, \begin{pmatrix} 8 \end{pmatrix}, \begin{pmatrix} 6 & 6 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 6 & 2 \\ 6 & 2 \end{pmatrix}$

**2.3**

$$\begin{pmatrix} 1-2i & 2+3i & 5+5i \\ 7+i & -3i & 3+2i \\ 7 & 10-4i & -3-3i \end{pmatrix} \quad \begin{pmatrix} 5+8i & -6-5i & 1-10i \\ -14-4i & 1+5i & -7+i \\ -15+i & -25+10i & -1+7i \end{pmatrix}$$

$$\begin{pmatrix} 11-3i & 27-12i & -2 \\ 2+3i & 10+8i & 7 \\ -10+2i & -12+12i & 2+9i \end{pmatrix} \quad \begin{pmatrix} -1 & 2-2i & -10-10i \\ 9+2i & -2+3i & 15+2i \\ 12-9i & -2-2i & 26+11i \end{pmatrix}$$

$$\begin{pmatrix} 13+3i & 29+2i & 2+2i \\ 2+5i & 2+10i & -7-2i \\ -10+4i & -16-2i & -2-9i \end{pmatrix}$$

**2.4**

(a) $$\begin{pmatrix} 2 & 3 & 1 \\ 3 & 3 & 1 \\ 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} \Rightarrow x = 2,\ y = -1,\ z = -2$$

(b) $$\begin{pmatrix} 1 & -2 & -1 \\ 3 & -5 & -2 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \\ 2 \end{pmatrix} \Rightarrow x = \frac{9}{2},\ y = -\frac{1}{2},\ z = \frac{11}{2}$$

**2.5**

(a) $\begin{pmatrix} 1 & 0 & 0 \\ -\frac{11}{5} & \frac{4}{5} & \frac{1}{5} \\ -\frac{1}{5} & -\frac{1}{5} & \frac{1}{5} \end{pmatrix}$ (b) $\begin{pmatrix} -17 & 23 & -\frac{5}{2} & -10 \\ 5 & -8 & \frac{1}{2} & 4 \\ 4 & -5 & \frac{1}{2} & 2 \\ -1 & 2 & 0 & -1 \end{pmatrix}$

**2.6**

(a) $E_{32}(1)E_3(\frac{1}{5})E_{23}(-1)E_{13}(-3)E_{12}(-2)$

(b) $E_{21}(-2)E_{31}(-3)E_{32}(1)E_{41}(-2)E_{42}(-2)E_{43}(-2)E_4(-1)E_3(\frac{1}{2})$
$\cdot E_{24}(-2)E_{23}(-2)E_{14}(-1)E_{13}(2)E_{12}(-1)$

**2.7** (a) 3 (b) 4

**2.8**

(a) $F_2 - F_1,\ F_4 - F_3 \Rightarrow$ dos filas iguales

(b) $F_2 - F_1,\ F_3 - F_2,\ F_4 - F_3$ y al resultado $F_3 - F_2,\ F_4 - F_3 \Rightarrow$ dos filas iguales

C 3

## SISTEMAS DE ECUACIONES LINEALES

**3.1** (a) $x_1 = 3,\ x_2 = 2,\ x_3 = 1$ (b) No tiene solución

**3.2** (a) $x = 3,\ y = 1,\ z = 1$ (b) $x = 1,\ y = 2,\ z = -2$

**3.3** $\text{rango}(A) = 3$

**3.4**

(a) $x_1 = 3 - 2x_3 - x_5,\ x_2 = -\frac{9}{2} + 3x_3 - \frac{1}{2}x_4 + 2x_5$

(b) No tiene solución

(c) $x_1 = -3 + 3x_3,\ x_2 = 2 - x_3$

(d) $x_1 = 5 - x_3 + x_4,\ x_2 = 2 - x_3 - x_4$

**3.5**

(a) $e_a = 0.1;\quad e_r = 0.02 \equiv 2\,\%$

(b) $e_a = 0.000015;\quad e_r = 0.000021287 \equiv 0.0021\,\%$

**3.6** Solución exacta: $x = \frac{13}{5}$, $y = -\frac{19}{5}$, $z = -5$.
Gauss sin pivoteo: $x = 2.598$, $y = -3.797$, $z = -5$. $e_a = 0.036$, $e_r = 0.0052 \equiv 0.59\,\%$.
Gauss con pivoteo parcial: $x = 2.585$, $x = -3.782$, $z = -4.979$. $e_a = 0.0314$, $e_r = 0.0046 \equiv 0.46\,\%$.

**3.7**

(a) $$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 2 & 0 & -1 & 1 \end{pmatrix};\quad U = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -6 \end{pmatrix}$$

(b) $$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 0 & 2 & \frac{1}{2} & 1 \end{pmatrix};\quad U = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

(c) $PA$ descompone en $L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$; $U = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{pmatrix}$

con $P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

**3.8** Solución exacta: $x_1 \approx 1203.84615$, $x_2 \approx -1153.84615$.
Solución aproximada: $x_1 = 1450$, $x_2 = -1500$. $e_a = 424,7519$, $e_r = 0.2547 \equiv 25\,\%$.

**3.9**

(a) $$\begin{cases} x_1^{(1)} = 0.3333,\ x_2^{(1)} = 0.0,\ x_3^{(1)} = 0.5714 \\ x_1^{(2)} = 0.1428,\ x_2^{(2)} = -0.3571,\ x_3^{(2)} = 0.4285 \\ x_1^{(3)} = 0.0714,\ x_2^{(3)} = -0.2142,\ x_3^{(3)} = 0.6632 \end{cases}$$

(b) $$\begin{cases} x_1^{(1)} = 0.9,\ x_2^{(1)} = 0.7,\ x_3^{(1)} = 0.6 \\ x_1^{(2)} = 0.97,\ x_2^{(2)} = 0.91,\ x_3^{(2)} = 0.74 \\ x_1^{(3)} = 0.991,\ x_2^{(3)} = 0.945,\ x_3^{(3)} = 0.782 \end{cases}$$

**3.10**

(a) $$\begin{cases} x_1^{(1)} = 0.3333,\ x_2^{(1)} = -0.1666,\ x_3^{(1)} = 0.5 \\ x_1^{(2)} = 0.1111,\ x_2^{(2)} = -0.2222,\ x_3^{(2)} = 0.6190 \\ x_1^{(3)} = 0.0529,\ x_2^{(3)} = -0.2328,\ x_3^{(3)} = 0.6485 \end{cases}$$

(b) $$\begin{cases} x_1^{(1)} = 0.9,\ x_2^{(1)} = 0.79,\ x_3^{(1)} = 0.758 \\ x_1^{(2)} = 0.979,\ x_2^{(2)} = 0.9495,\ x_3^{(2)} = 0.7899 \\ x_1^{(3)} = 0.9949,\ x_2^{(3)} = 0.9574,\ x_3^{(3)} = 0.7914 \end{cases}$$

C 4

## ESPACIOS VECTORIALES

**4.1**

(a) Independientes

(b) Dependientes: $(1,2,3)-(1,3,2)-(0,-1,1)$

(c) Dependientes: $0\cdot(1,0,1,0)+(2,1,3,1)+(0,1,1,1)-(2,2,4,2)$

**4.2** (a) 3 (b) 2 (c) 2

**4.3** (a) Independientes (b) Independientes (c) Dependientes

**4.4** (a) Forman base (b) Forman base (c) No forman base

**4.5**

(a) $\text{rango}(A)=2$ y $\text{rango}(C)=3$, luego contiene una base

(b) $B=\{(1,5,1),(2,1,0),(1,0,0)\}$

**4.6** Como $\text{rango}(S)=3$, es sistema generador. $S_1=\{x+1,x-1,x^2-1\}$

**4.7**

(a) $\text{rango}(\mathcal{B}')=3$

(b) $(3,-5,7)_{\mathcal{B}'}$

(c) $\mathbf{x}_{\mathcal{B}'}=\begin{pmatrix}1&0&0\\-1&1&0\\0&-1&1\end{pmatrix}\mathbf{x}_{\mathcal{B}}$

**4.8** (a) Sí (b) No (c) Sí (d) Sí

**4.9** No

**4.10** Sí

**4.11**

(a) Dimensión: 2; base: $\{(-1,3,2,0),(1,0,0,1)\}$

(b) Dimensión: 2; base: $\{(-1,1,0,0),(7,0,-3,1)\}$

(c) Dimensión: 2; base: $\{(-1,-1,1,0),(1,-1,0,1)\}$

**4.12** (a) $\begin{cases}x_1-x_3=0\\x_1-x_2+x_4=0\\x_1-x_2-x_5=0\end{cases}$ (b) $\begin{cases}-6x_1+2x_3-4x_4=0\\x_1-x_5=0\end{cases}$

C 5

## APLICACIONES LINEALES

**5.1** Todas son lineales.

**5.2**

(a) $\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & -3 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & -1 & 1 \end{pmatrix}$

**5.3** Base de $\mathcal{M}_{3\times 2}(\mathbb{R})$:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \right\}$$

Matriz de la aplicación traza: $T = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

**5.4** $T = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 01 & \\ 0 & 0 & 1 & 0 \end{pmatrix}$

**5.5**

(a)

$$(f+f')\circ g \to \begin{pmatrix} 26 & 7 \\ -1 & 1 \end{pmatrix} \quad (f+f')\circ g' \to \begin{pmatrix} 7 & 27 \\ 1 & -2 \end{pmatrix}$$

$$f\circ(g+g') \to \begin{pmatrix} 12 & 11 \\ 6 & 6 \end{pmatrix} \quad g\circ(f+f') \to \begin{pmatrix} 5 & -1 & 7 \\ 7 & 1 & 14 \\ 9 & 3 & 21 \end{pmatrix}$$

$$g'\circ(f+f') \to \begin{pmatrix} 7 & 1 & 14 \\ 6 & -2 & 7 \\ 3 & -3 & 0 \end{pmatrix} \quad (g+g')\circ f' \to \begin{pmatrix} 9 & 3 & 3 \\ 11 & 2 & 0 \\ 9 & 3 & 3 \end{pmatrix}$$

$$(g+g')\circ(f+f') \to \begin{pmatrix} 12 & 0 & 21 \\ 13 & -1 & 21 \\ 12 & 0 & 21 \end{pmatrix} \quad (f+f')\circ(g+g') \to \begin{pmatrix} 33 & 34 \\ 0 & -1 \end{pmatrix}$$

(b) $(f+f')\circ g$: Biyectiva; $(f+f')\circ g'$: Biyectiva; $f\circ(g+g')$: Biyectiva; $g\circ(f+f')$: Nada; $g'\circ(f+f')$: Nada; $(g+g')\circ f$': Nada; $(g+g')\circ(f+f')$: Nada; $(f+f')\circ(g+g')$: Biyectiva.

**5.6** La matriz de $f$ es $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ y la de $g$ es $B = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. Basta ver que $AB = I$.

**5.7** Es invertible porque $\det(T) \neq 0$. $T^{-1} = \begin{pmatrix} -2 & -2 & 3 \\ 0 & -1 & 1 \\ 3 & 4 & -5 \end{pmatrix}$

**5.8** $\begin{pmatrix} -6 & -10 & -\frac{25}{2} \\ -5 & -3 & -\frac{7}{2} \\ 6 & 6 & 8 \end{pmatrix}$

**5.9** $\begin{pmatrix} 4 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}$

**5.10**

(a) Base de $\ker(A) = \{(-1,0,1)\}$. $\mathrm{Im}(A) : \{x+y-z=0$
$A$ no es ni inyectiva ni sobreyectiva.

(b) $\ker(B) = \{\mathbf{0}\}$, $\mathrm{Im}(B) = \mathbb{C}^2$. $B$ es inyectiva y sobreyectiva.

(c) Base de $\ker(D) = \{(2i,-2,1)\}$. $\mathrm{Im}(D) : \{-ix+z=0$
$D$ no es ni inyectiva ni sobreyectiva.

C 6

## DIAGONALIZACIÓN

### 6.1

(a) $p(\lambda) = \lambda^2 + \lambda - 2$; Autovalores: $-2$, $1$.

(b) $p(\lambda) = \lambda^2 - 6\lambda + 9$; Autovalores: $3$ (doble).

(c) $p(\lambda) = -\lambda^3 + 6\lambda^2 - 11\lambda + 6$; Autovalores: $1$, $2$, $3$.

(d) $p(\lambda) = -\lambda^3 - \lambda^2 + \lambda + 1$; Autovalores: $-1$ (doble), $1$.

(e) $p(\lambda) = -\lambda^3 - 10\lambda^2 - 32\lambda - 32$; Autovalores: $-4$ (doble), $-2$.

(f) $p(\lambda) = -\lambda^3 + \lambda^2$; Autovalores: $0$ (doble), $1$.

(g) $p(\lambda) = \lambda^4 - 2\lambda^3 + \lambda^2$; Autovalores $0$ (doble), $1$ (doble).

(h) $p(\lambda) = \lambda^4 - 4\lambda^3 + 3\lambda^2 + 4\lambda - 4$; Autovalores $2$ (doble), $-1$, $1$.

### 6.2

(a) $\{(-2, 1), (-1, 1)\}$.

(b) No hay base de autovectores.

(c) $\{(1, -2, 2), (1, -3, 3), (-1, 0, 1)\}$.

(d) No hay base de autovectores.

(e) $\{(1, 2, 0), (1, 0, 6), (2, 3, 1)\}$

(f) No hay base de autovectores.

(g) No hay base de autovectores.

(h) No hay base de autovectores.

### 6.3

(a) $D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{pmatrix} \quad P = \begin{pmatrix} 0 & 2 & -1 \\ 1 & 1 & 3 \\ 3 & 1 & 3 \end{pmatrix}$

(b) No es diagonalizable en $\mathbb{R}^3$.

(c) $D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

(d) $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad P = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$

(e) No es diagonalizable en $\mathbb{R}^3$.

**6.4**

(a) $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \quad P = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}$

(b) $J = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad P = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & 3 \\ 1 & 0 & 4 \end{pmatrix}$

(c) $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \quad P = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}$

(d) $J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad P = \begin{pmatrix} 1 & 1 & 3 \\ 4 & 0 & 0 \\ 3 & 0 & 1 \end{pmatrix}$

**6.5**

(a) Autovalores: 0 y $2-i$; Autovectores: $(1+i,-1)$ y $(i,1)$. Matrices de paso:

$$\begin{pmatrix} 1+i & i \\ -1 & 1 \end{pmatrix}$$

No hay matriz de paso real.

(b) Autovalores: $1-3i$ y $1+3i$; Autovectores: $(-i,3)$ y $(i,3)$. Matrices de paso compleja y real:

$$\begin{pmatrix} -i & i \\ 3 & 3 \end{pmatrix} \qquad \begin{pmatrix} 0 & -1 \\ 3 & 0 \end{pmatrix}$$

(c) $i$ y $-i$ (dobles); Autovectores: $(-i,0,1,0)$, $(i,-i,0,1)$, $(i,0,1,0)$, $(-i,i,0,1)$. Matrices de paso compleja y real:

$$\begin{pmatrix} -i & i & i & -i \\ 0 & -i & 0 & i \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

**6.6** $$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

C 7

## ECUACIONES LINEALES EN DIFERENCIAS

**7.1**

(a) $x_k = \frac{5}{6}2^k + \frac{5}{3}(-1)^k$

(b) $x_k = 2^{k-1}$

(c) $x_k = c_1 3^k + c_2 k 3^k + c_3(-1)^k$

(d) $x_k = c_1 \cos\left(k\frac{\pi}{2}\right) + c_2 \operatorname{sen}\left(k\frac{\pi}{2}\right) + c_3 k \cos\left(k\frac{\pi}{2}\right) + c_4 \operatorname{sen}\left(k\frac{\pi}{2}\right)$

**7.2** $u_n = 6^n$, $v_n = 6^n$

**7.3** $x_n = \frac{\sqrt{5}}{5}\left(\frac{1}{2} + \frac{\sqrt{5}}{2}\right)^n - \frac{\sqrt{5}}{5}\left(\frac{1}{2} - \frac{\sqrt{5}}{2}\right)^n$

**7.4** $c\begin{pmatrix} 3 \\ 2 \end{pmatrix}$

**7.5** $c\begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$

C 8

## ESPACIO VECTORIAL EUCLÍDEO

**8.1** $\begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

**8.2** $\{(1,0,0),(0,-2,0),(0,0,5)\}$

**8.3** $\{(-2,1,0),(-3,-6,5)\}$

**8.4**

(a) No

(b) Sí

(c) No

**8.5** $\{-5x^3+3x, -15x^2+12x+9\}$

**8.6** $\mathbf{c} = \left(-\frac{7}{17}, -\frac{6}{17}, -\frac{8}{17}, \frac{15}{17}, 0\right) + \left(\frac{7}{17}, \frac{6}{17}, \frac{25}{17}, \frac{19}{17}, 1\right)$

**8.7** $(2,1,1)$

**8.8**

(a) $\{x, x^2-x\}$

(b) $\left(3+\frac{9}{e}\right)x^2 - \left(\frac{8}{e}+2\right)x$

**8.9**

(a) $(1,-1)$

(b) $(2,-2,1)$

C 9

## ESPACIO AFÍN

**9.1** Cambio de sistema de referencia:

$$\begin{pmatrix} 1 \\ \hline x \\ y \end{pmatrix} = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 1 & 2 & -1 \\ 2 & 2 & 2 \end{array}\right) \begin{pmatrix} 1 \\ \hline x' \\ y' \end{pmatrix}$$

Ecuación de la recta: $4x' + y' - 2 = 0$

**9.2**

(a) $(x, y, z) = (1, 2, -1) + t(1, -1, 4)$

(b) $(x, y, z, t) = (0, 1, -1, 0) + t(0, 1, -1, 1)$

(c) $(x, y, z) = t(-1, 2, 3)$

**9.3** Se cruzan.

**9.4**

(a) $x - 2y + z = 1$

(b) $5x - 7y + z = -3$

(c) $-x + 5y + 2z = 0$

**9.5** $d(P, r) = \sqrt{5}, \quad d(P, \pi) = \frac{2}{\sqrt{11}}$

**9.6** $d(r_1, r_2) = d\left((2, 1, -1, 0), \left(\frac{3}{2}, 0, \frac{1}{2}, 1\right)\right) = \frac{3}{\sqrt{2}}$

**9.7** (a) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ 0 & \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & -1 \end{pmatrix}$

**9.8**

(a) $\left(\frac{2}{3}, \frac{5}{3}, -\frac{4}{3}\right)$

(b) $\left(\frac{1}{3}, -\frac{7}{3}, -\frac{5}{3}\right)$

(c) $\left(\frac{8}{7}, \frac{11}{7}, -\frac{3}{7}, \frac{10}{7}\right)$

# Índice terminológico


**A**

aplicación, 378
- afín, 363
- biyectiva, 202, 213, 378
- composición, 201, 378
- diagonalizable, 231, 240
- dominio, 378
- identidad, 192
- imagen, 209, 211, 213, 378
- inversa, 202, 378
- inyectiva, 210, 213, 378
- lineal, 191
- núcleo, 209, 210, 213
- nula, 192
- producto por escalar, 200
- sobreyectiva, 213, 378
- suma, 200

aritmética de coma flotante, 103
automorfismo, 193
autovalor, 229, 234, 303
autovector, 229, 234, 303

**B**

base
- canónica, 150
- definición, 149
- ortonormal, 324

binomio de Newton, 381
- para matrices, 269

**C**

cambio de base
- en aplicaciones lineales, 204
- en espacios vectoriales, 158

cambio de sistema de referencia, 351
clase de equivalencia, 377
combinación lineal, 140, 147, 191
complemento ortogonal, 326
conjunto, 375
- cociente, 182, 377
- diferencia, 377
- intersección, 376
- partes de, 376
- unión, 376
- vacío, 376

coordenadas, 150
cuerpo, 15, 136, 183, 380

**D**

dependencia lineal, *véase* linealmente independiente, 152
desigualdad
- de Bunyakowski, 314
- de Cauchy, 314
- de Schwarz, 313, 348
- triangular, 16, 314

determinante, 59
- de Vaandermonde, 86
- propiedades, 60–66

dimensión, 153
dirección, 353
distancia, 352
doble implicación, 375

**E**

ecuación
- característica, 294
- en diferencias, 286

ecuaciones diferenciales, 124
elemento, 375
endomorfismo, 192, 227
equivalencia, *véase* doble implicación
error
- absoluto, 103
- de proyección, 330, 334
- de redondeo, 102, 397

de truncamiento, 102
relativo, 103
escalares, 35
espacio
afín, 349
espacio afín, 349
espacio métrico, 352
espacio unitario, 312
espacio vectorial
de dimensión finita, 156
de dimensión infinita, 156
definición, 136
euclídeo, 311
estructura algebraica, 379

**F**
fórmula
de de Moivre, 20
de Euler, 19
de la dimensión
en aplicaciones, 213
para subespacios, 175
factorización LU, 107
forma canónica de Jordan, 256
real, 265
fractal, 26
conjunto de Julia, 34
curva de Koch, 27
de Mandelbrot, 27, 30
función, *véase* aplicación

**H**
homomorfismo, 192
homotecia, 372

**I**
imagen
de un subespacio, 193
de una base, 194
implicación, 375
inclusión, 376
isomorfismo, 192

**L**
linealmente independiente, 141, 142
longitud, *véase* norma vectorial

**M**
método
de Gauss, 49, 52, 54, 72, 80, 92–94, 103, 108
con pivoteo parcial, 105
con pivoteo total, 107
de Gauss-Seidel, 117
de Jacobi, 116, 122
de ortogonalización de Gram-Schmidt, 319
del orlado, 70
matriz
adjunta, 74
adjunta de un elemento, 58
antisimétrica, 88
banda, 127
columna, 36
cuadrada, 37
de adyacencia, 80
de cambio de base, 158
de cofactores, *véase* matriz adjunta
de Gram, *véase* matriz de un producto escalar
de Jordan, 253
de paso, 243, 256
de permutación, 112
de proyección, 333
de un producto escalar, 316
de un sistema, 48
de una aplicación afín, 364
de una aplicación lineal, 195, 200, 227
definición, 36
diagonal, 227, 268
por cajas, 253, 257
diagonal dominante, 119
diagonalizable, 231
elemental, 50, 57, 65, 108
elemental de Jordan, 252
equivalente, 252
fila, 36
identidad, 42
inversa, 45, 54, 74
invertible, *véase* matriz regular, 57
menor, 69
número de condición, 114
nilpotente, 88, 269
nula, 39
ortogonal, 324
producto, 40, 43, 46

producto por escalar, 38
regular, 45, 65, 88
simétrica, 37, 88
singular, 45
submatriz, 58
suma, 38
traspuesta, 37, 44, 46
tridiagonal, 127
vacía, 115, 127
mejor aproximación, 335
multiplicidad
algebraica, 239
geométrica, 239

**N**
número complejo
argumento, 17
conjugado, 13
definición, 12
división, 14, 19
forma binomial, 12
forma polar, 19
forma trigonométrica, 18
imaginario puro, 12
módulo, 16
multiplicación, 13, 19
parte imaginaria, 12
parte real, 12
potencia, 20
raíz, 20
representación gráfica, 15, 18
resta, 13
suma, 13
unidad imaginaria, 12
norma
euclídea, 113
matricial, 114
vectorial, 113, 313

**O**
origen de coordenadas, 351
ortogonalidad, 317, 325

**P**
partición de multiplicidad, 255
plano complejo, 15
polinomio característico, 235
potencia de una matriz, 268
principio de inducción, 61–63, 233, 380
producto
escalar, 311
interno, 312
proyección ortogonal, 329, 358
Python
*slicing*, 78, 389
`E`, 25
`I`, 24
`LUdecomposition`, 121
`Matrix`, 77, 215
`Symbol`, 25
`_`, 23
`abs`, 24
`append`, 220
`arg`, 24
`astype`, 395
`bmat`, 369
`break`, 31
`col_joint`, 369
`concatenate`, 369
`conjugate`, 24
`diagflat`, 123
`diagonal`, 123
`diag`, 128
`dot`, 180, 340
`dtype`, 395
`eigenvals`, 307
`eigenvects`, 275, 306
`else`, 31
`evalf`, 24
`exit`, 123
`expand`, 25
`eye`, 271
`float`, 31
`for`, 31, 390
`for-else`, 31
`if`, 123, 390
`import`, 391
`integrate`, 341
`lambda`, 129
`linalg.cond`, 122
`linalg.eigvals`, 275
`linalg.eig`, 275
`linalg.lstsq`, 181, 342
`linalg.solve`, 119
`linspace`, 128
`lu`, 121
`matplotlib`, 28, 31, 129

matrix, 76, 215
norm, 341
nullspace, 214
ones, 128
path, 182
r_, 369
range, 389
raw_input, 221, 385
reshape, 180
return, 393
re, 25
roots, 271
row_join, 273, 369
rref, 79, 120
shape, 77
simplify, 23, 397
solve_lineal_system, 120
solve, 25, 307
sqrt, 24
subs, 307
symbols, 271, 341, 397
type, 387
determinante, 79
diccionario, 217
función, 127, 393
argumento por defecto, 123
objetos, 394
operaciones con complejos, 23
rango, 181
str, 220
tipos de datos, 386
cadenas, 388
complejos, 387
enteros, 387
listas, 388
reales, 387

**R**

raíces de la unidad, 21
rango
de un conjunto de vectores, 146
de una matriz, 69, 100, 142, 213
regla de Cramer, 95
relación
de equivalencia, 182, 377
de orden, 15, 377
rotación, 197, 230, 237

**S**

símbolo de Kronecker, 42
simetría, 197, 365
sistema de ecuaciones lineal
bajada, 108
compatible, 91
condicionamiento, 113, 114
definición, 47
determinado, 91
grado de libertad, 93
homogéneo, 94, 165, 170
incompatible, 91
indeterminado, 91
métodos directos, 103
métodos iterativos, 115
resolución, 58
solución, 48
solución aproximada, 337
solución trivial, 94
subida, 49, 107
triangular, 49
sistema de referencia, 351
sistema generador, 147
de un subespacio, 163
sistemas dinámicos discretos, 286
subconjunto, 376
subespacio vectorial, 162, 170
ecuaciones implícitas, 171
intersección, 174
invariante, 229, 254
máximo, 254
propio, 239
propio generalizado, 253
suma, 174
suma directa, 176

**T**

teoría de grafos, 80
teorema
de Jordan, 253
de Kronecker-Capelli, *véase* teorema de Rouché-Frobenius
de Pitágoras, 318, 348
de Rouché-Fontené, *véase* teorema de Rouché-Frobenius
de Rouché-Frobenius, 98, 144, 337
fundamental del Álgebra, 22
traslación, 364
traza de una matriz, 189, 222

**V**

valor propio, *véase* autovalor
valores singulares, 342
variedad
  afín, 353
  lineal, *véase* subespacio vectorial
vector, 137
  de equilibrio, *véase* vector estacionario
  estacionario, 302
  nulo, 140
  propio, *véase* autovector

# Índice de autores

**A**
Al-Jwarizmi, Muhammad ibn Musa, (˜780–850), 3
Argand, Jean-Robert (1768–1822), 15

**B**
Bellavitis, Giusto (1803–1880), 135
Bessel, Friedrich Wilhem (1784–1846), 348
Bombelli, Rafael (1526–1572), 13
Bunyakowski, Viktor (1804–1889), 314

**C**
Capelli, Alfredo (1855–1910), 98
Cardano, Gerolamo (1501–1576), 11, 58
Cauchy, Augustin Louis (1789–1857), 58, 228, 314, 320
Cayley, Arthur (1821–1895), 38
Cramer, Gabriel (1704–1752), 95

**D**
d'Alambert, Jean Le Rond (1717–1783), 22
de Moivre, Abraham (1667–1754), 20
Doolittle, Myrick Hascall (1830–1911), 107

**E**
Euler, Leonhard (1707–1783), 12, 19, 80, 228, 349

**F**
Fibonacci, *véase* Pisa, Leonardo da (1170–1250)
Fontené, Georges (1848–1923), 98
Fourier, Joseph (1768–1830), 336
Frobenius, Ferdinand Georg (1849–1917), 69, 98

**G**
Gauss, Carl Friedrich (1777–1855), 11, 12, 22
Girard, Albert (1595–1632), 22
Gram, Jørgen Pedersen (1850–1916), 316
Grassmann, Hermann (1809–1877), 135, 140, 191

**H**
Hamilton, William Rowan (1805–1865), 135
Hilbert, David (1862–1943), 228

**J**
Jacobi, Gustav Jakov (1804–1851), 116
Jordan, Camille (1838–1922), 245
Julia, Gaston (1893–1978), 34

**K**
Koch, Helge von (1870–1924), 26
Kronecker, Leopold (1823–1891), 42, 98

**L**
Lagrange, Joseph-Louis (1736–1813), 228, 235
Laplace, Pierre-Simon (1749–1827), 320
Leibniz, Gottfried Wilhelm von (1646–1716), 58
Lorenz, Edward (1917–2008), 27

**M**
Maclaurin, Colin (1698–1746), 95
Mandelbrot, Benoît (1924–2010), 26, 27
Markov, Andréi Andréyevich (1856–1922), 300

**N**
Newton, Isaac (1643–1727), 381

**P**

Parseval, Marc-Antoine (1755–1836, 348
Peano, Giuseppe (1858–1932), 135
Pisa, Leonardo da (1170–1250), 286
Poincaré, Henri (1854–1912), 27

**R**

Rey Pastor, Julio (1888–1962), 98
Rouché, Eugène (1832–1910), 98

**S**

Sarrus, Pierre Fréderic (1798–1861), 60
Schmidt, Erhard (1876–1959), 320
Schwarz, Hermann Amandus (1843–1921), 313
Seidel, Philipp Ludwig von (1821–1896), 117
Seki, Kowa (1642–1708), 58
Steinitz, Ernst(1871–1928), 140
Sylvester, James Joseph (1814–1897), 36

**T**

Turing, Alan (1912–1954), 113

**V**

Vandermonde, Alexandre (1735–1796), 86

**W**

Wessel, Caspar (1745–1818), 15
Weyl, Hermann Klaus Hugo (1885–1955), 313

# Bibliografía

La mayoría de libros publicados que tratan sobre álgebra lineal incluyen buena parte de los contenidos cubiertos en este texto, por lo que además de los citados aquí [3, 6, 7, 9, 11, 12, 17, 22], el lector puede encontrar material similar en muchos otros textos. Por otra parte, internet se ha convertido en una fuente inagotable de recursos, tanto de páginas web como de material sin publicar, que está a disposición de todos gracias al trabajo de personas de todo el mundo; algunos de esos materiales son citados aquí [1, 2, 5, 8, 10, 13, 18, 23], aunque son tantas las páginas visitadas durante la elaboración de este texto que a veces resulta difícil referenciar todo el material empleado.

---

## REFERENCIAS


[1] The mactutor history of mathematics archive. URL: `www.gap-system.org/~history/`.

[2] Wikipedia. URL: `www.wikipedia.org/`.

[3] J. Acher. *Álgebra Lineal y Programación Lineal*. Ed. Montaner y Simón, 1967.

[4] I. Arregui, A. Ferreiro, y J.A. García. Programación en python orientado a la ingeniería. II Curso intensivo i-MATH de Software Libre orientado a Ciencias e Ingeniería, 2009.

[5] Rafael López Camino. Aplicaciones afines. Material docente para el alumno. URL: `www.ugr.es/~rcamino/docencia/geo1-03/g1tema8.pdf`.

[6] Agustín de la Villa. *Problemas de álgebra*. CLAGSA, 1994.

[7] L.I. Goloviná. *Álgebra lineal y algunas de sus aplicaciones*. Editorial Mir, 1980.

[8] Chelo Ferreira González. Modelos discretos elementales. ecuaciones en diferencias. URL: `pcmap.unizar.es/~chelo/teoria/docum_teor_out/tema4.pdf`.

[9] Stanley I. Grossman. *Álgebra lineal*. McGraw-Hill, 1996.

[10] Jim Hefferon. Linear algebra. URL: joshua.smcvt.edu/linearalgebra/.

[11] Bernard Kolman. *Álgebra lineal con aplicaciones y MATLAB*. Prentice Hall, 1999.

[12] Ron Larson, Bruce H. Edwards, y David C. Falvo. *Álgebra lineal*. Ediciones Pirámide, 2004.

[13] Vicent Aranu Llombart. Ampliación de estructura de computadores. URL: www.uv.es/varnau/AEC_520.pdf.

[14] Raúl González Luque. Python para todos. URL: mundogeek.net/tutorial-python/.

[15] Deivi Luzardo y Alirio J. Peña. Historia del Álgebra lineal hasta los albores del siglo xx. *Divulgaciones Matemáticas*, 14(2):153–170, 2006.

[16] Andrés Marzal y Isabel Gracia. *Introducción a la programación con Python*, volume 147. 2002.

[17] Luis Merino y Evangelina Santos. *Álgebra Lineal con métodos elementales*. Thomson, 2006.

[18] Rafel Amer Ramon y Vicenç Sales i Inglès. Àlgebra lineal. problemes resolts. URL: ruth.upc.es/assignatures/algebra.html.

[19] Guillermo Dávila Rascón. El desarrollo del álgebra moderna. parte i: El álgebra en la antigüedad. *Apuntes de Historia de las Matemáticas*, 1(3):5–21, 2002.

[20] Guillermo Dávila Rascón. El desarrollo del álgebra moderna. parte ii: El álgebra de las ecuaciones. *Apuntes de Historia de las Matemáticas*, 2(1):27–58, 2003.

[21] Guillermo Dávila Rascón. El desarrollo del álgebra moderna. parte iii: El surgimiento del álgebra abstracta. *Apuntes de Historia de las Matemáticas*, 2(2):38–78, 2003.

[22] Jesus Rojo. *Algebra Lineal*. McGraw-Hill, 2001.

[23] Luis Rández. Eliminación gaussiana con pivote parcial. URL: pcmap.unizar.es/~pilar/pivote.pdf.

[24] Ferrán Mir Sabaté. El teorema fundamental del álgebra. 2005. URL: http://personal.telefonica.terra.es/web/mir/ferran/id1.html.

[25] Gema Calbo Sanjuán y Juan C. Cortés López. Aplicación de las matrices invertibles en criptografía. *Ensayos: Revista de la Facultad de Educación de Albacete*, 18:279–284, 2003.